The fishPi project
Regional co-operation in fisheries data collection

STRENGTHENING REGIONAL COOPERATION IN FISHERIES DATA COLLECTION

Report to the European Commission in Fulfilment of Grant Award: EU MARE/2014/19

June 2016

MARE/2014/19

Project funded by the European Union

Published by	To be agreed
This report is available on	To be agreed
Dissemination Statement	Draft for internal review only until approved by the funder.

This report was prepared for

DG MARE
European Commission
Directorate-General for Maritime Affairs and Fisheries
B-1049 Brussels

The report was compiled by

MASTS
SOI-Gatty Laboratory
University of St Andrews
East Sands
St Andrews
Fife KY16 8LB
United Kingdom

Contact Dr M. A. James
Telephone +44(0)1334467312
Email maj8@st-andrews.ac.uk

Revision Record

Revision Record	Issue Date	Revision Details
Draft Final Report	06/06/2016	Submitted

Project funded by the European Union

Acknowledgements

We acknowledge the EU Commission together with the partner institutions involved in this project (EU MARE 2014/19) for their financial support, without which this project would not have been possible. We are grateful to the EU MARE team led by Venetia Kostopoulou for their helpful comments and support throughout the project. The project is particularly indebted to those Member States who's positive and enlightened approach to data sharing enabled us to make important progress towards achieving the objectives of the project. The high response rate and positive outcome of the consultation conducted at the end of the project bears testament to the potential value of this work and we are grateful to all consultees for their contribution. Where appropriate, members of the project team liaised closely with colleagues undertaking a parallel project in the Mediterranean and Black Sea regions and we thank the Project Co-ordinator Mariateresa Spedicato for facilitating this process.

Finally, we would like to thanks the many individuals who contributed to the completion of this project, in particular the work package leaders and case study leaders who took on the added responsibility of delivering substantial sections of the project, and the core teams involved, whose hard work, expertise and dedication produced the bulk of the work. Also the external experts for their advice and guidance and the MASTS team for allowing the scientists to concentrate on delivering the science. The full list of fishPi participants can be found at the end of the report.

Table of Contents

fishPi Executive Summary

fishPi was a research project with the aim of "Strengthening regional cooperation in the area of fisheries data collection". The project brought together over 40 experts from 13 scientific institutes in 12 countries (10 member states (MS)) and two internationally recognised survey design experts. It was funded by EU MARE grant MARE/2014/19, with a 14 month timeline commencing in April 2015.

This project has trialled the way sampling designs would be developed in a regional setting and showed that collaboration and consultation is required at face to face meetings through regional groups that focus on a particular group of fisheries. The project was the first step in this process and one of the main outcomes is the framework to take the process forward; developing data formats, data sharing agreements and easily accessible software for data sharing, checking and analysis, and for the simulation testing of sampling designs. These designs are predicated on common data collection protocols and the use of the appropriate statistical estimators; the implementation of such designs would thus require the adoption of the standard survey sampling techniques and the use of common sampling and estimation routines by the sampling institutions.

The main findings of four commercial fishery case studies were that considerable improvements can be made by adopting regional designs, by which we mean the adoption of a common metric used for stratification (such as port size or fleet segments etc), though with the nation being retained as a level of stratification within the overall design. Such designs would potentially provide unbiased and more precise estimates than the coordinated national data collection schemes operating at present. The main issues found in the operation of national sampling designs at present is the incomplete sampling coverage of the regional population, and that the allocation of sampling effort unilaterally at national level does not represent the best use of the available resource.

Further work needs to be carried out in identifying appropriate fisheries for regional sampling, testing the assumptions of biological data collection, species selection protocols and sampling effort in relation to the data needs of end users.

Small scale and recreational fisheries, by-catch and stomach content sampling programmes do not have established regional sampling schemes, and data collection is not routinely carried out by MS to the same extent as the existing commercial fisheries sampling. Through end-user consultations and case studies, this project has shown that these data have particular requirements. There are other end-users usually distinct from those of the main end users of commercial fisheries data to be considered. As a consequence, data needs and the potential for regional cooperation should be carefully explored with end-users as a prerequisite to the design stage of any regional scheme. There could be considerable resource implications, both technical and economic, for implementation.

A major remit of the project was to develop guidelines to evaluate the quality of data at national and regional levels using shared tools. To that end an R library has been developed and made available on a public access website (https://github.com/ldbk/fishPifct). The data structure upon which the work was developed was an upgrade of the data exchange format for the current regional database. The suite of quality checks functions have been designed to provide considerable flexibility in their use as it is recognized that the formats and reference lists will evolve over time, under the RCG umbrella. The ideal time frame for these checks would be quarterly checks at a national level for submission to a regional data base at the end of February. Regional quality checks can then be carried out in March. This time frame might need discussion and adaptation before being adopted at a regional level. The functions developed during the project are to be seen as a first version aimed at continuous improvement for the benefits of all the Member States. Details of a short term development plan and resources needed are proposed for a gradual implementation.

A review of the historical operation of RCM has highlighted improvements in regional coordination between MS. The establishment of the fishery activity matrix and the regional data base are considered to be the main elements that has led to a common understanding of regional fisheries and resulted in harmonised codes for métiers, species, harbours and areas. It has emphasised the building of links, trust, skills and understanding of experts across MS and the important contribution specific projects (COST and WebGR) have made to that process. However in order to gain the full potential benefits from regional cooperation there is a clear need to develop regional work over longer time frames than hitherto, and with appropriate funding mechanisms and organisational infrastructures.

The major findings of fishPi formed the basis of a written consultation addressed to all National Correspondents, the Heads of Institutes from Member States attending the RCM NS&EA, RCM NA and RCM Baltic, and ICES and EU, who are considered the main end-users of the data. Responses were received from 15 out of 17 Member States. The conclusions from the consultation included strong support for the overall concept of regional sampling designs using probability-based selection methods and associated regional estimation methods, and strong support for the use of the regional database and the operation of data quality indicators. Respondents also commented that national sampling requirements need to be taken into consideration and that additional resources may be required.

The added value brought by the fishPi project has been substantial. Most noticeably this can be seen in the establishment of co-operative working relationships between experts in the scientific institutes involved and the extent to which all participants have broadened their skills base and understanding of the statistical principles of design based sampling and probability based selection.

fishPi Recommendations

Regional sampling designs have the potential to improve the statistical validity, data quality and cost-effectiveness of data collection and are directly related to the end-user needs. This applies equally to large scale commercial, small scale, and recreational fisheries, bycatch sampling and ecosystem based data collection. Therefore steps should be taken to develop and implement regional sampling designs, taking into account that the primary need is to identify the fisheries most suitable for such regional designs. It is the opinion of the fishPi participants that all the case study fisheries have the potential to be viable regional data collection schemes.

Specific regional sampling groups should be established for these key fisheries to oversee the development of regional sampling schemes. These groups should be inclusive; with data collectors, lead scientists in the countries involved, and experts in sampling, all in attendance. They should work in consultation with the main end-users. These groups should operate along the lines of ICES expert groups under the umbrella of the Regional Coordination Groups (RCGs).

A mechanism needs to be found to respect the autonomy of action of Member States and scientific institutions, in the collecting of data for national use, and reconcile this with requirements to collect data for regional data needs, in accordance with sampling protocols and data quality criteria set at the regional level.

Data flows, data exchange formats, data quality checking and a regional data base linked to national data sources are an integral part of regional sampling coordination and the production of regional estimates and reporting. Standardized data quality checks should be implemented on national and regional data, following an agreed timetable. A regional data database, WebGR and the development of software tools in R, need to have secure long term funding to meet developing needs for improvement.

The RCG needs to be empowered with suitable autonomy of action and funding mechanisms to be able to facilitate the appropriate use of expertise, and staff time, needed for the development and implementation of regional sampling plans.

Introduction

fishPi was a collaborative research project in response to the European Commission’s call for proposals MARE/2014/19, “Strengthening regional cooperation in fisheries data collection”.

The fishPi consortium partners included 13 scientific institutions, from 10 EU Member States (MS). In total over 40 individuals were involved in the work of the project, with skills ranging from fisheries science, survey sampling, and statistics, to software development, IT and management. Two external statistical experts also advised on the project work. The consortium included individuals attending the Regional Coordination Meetings (RCM) for the North Sea and Eastern Arctic (RCM NS&EA), the North Atlantic (RCM NA) and the Baltic (RCM Baltic), however to meet the specification set out in the call the proposal was nominally from members states involved in the RCM NS&EA.

The aims of the project were very broadly to devise sampling designs for regional fisheries; large scale commercial fisheries, small scale and artisanal fisheries, recreational fisheries, designs for the collection of data on the incidental by-catch (of marine mammals, birds, reptiles and rare fish), and designs for stomach contents data to support work on trophic interactions. The commercial fishery designs were addressed by means of case studies focussing on fisheries exploiting shared stocks in regional seas. An overriding aim of the project was to set out designs based on probability based selection protocols, and use the appropriate estimation methods, with the clear intention of improving the statistically validity of fisheries data collection.

To compliment these designs the project also had a major remit to develop guidelines and provide common software tools to evaluate the quality of fisheries data at both the national and regional level. Here the intention was again to promote commonly accepted quality assurance (QA) standards. The use of a standard sampling data format and harmonised code lists, compatible with the use of the regional data base (RDB), enables region-wide transparent and consistent data standards. A calendar for the application of these data quality routines was also developed.

These regional sampling designs and quality indicators were set in the wider context in which regional sampling was to be undertaken. This involved a review of the existing coordination mechanisms, principally the historic operation of RCMs to access the strengths and the weaknesses of the existing arrangements. This, together with the results of the sampling designs, informed the drafting of a regional work plan outlining the steps needed for the implementation of regional designs. Finally the designs, QA procedures and work plan, were the subject of a consultation process which was conducted across all MS within the region.

The project was organised as a series of four linked work packages (WP) operating under a project management committee (PMC).

- WP1 would address the review of existing arrangements, the regional work plan and the consultation.
- WP2 would undertake the collation of a regional data set for each of four commercial case study fisheries, and run simulation models to test regional designs. It would also prepare sampling and estimation guidelines and define the format and code lists needed for regional data collection.

- WP3 would undertake reviews, end user consultations and case studies to address the needs of small scale recreational sampling designs, designs for by-catch and designs for stomach sampling.
- WP4 would address the QA indicators for data, write software tools to provide such checks and propose a calendar for their implementation.
- PMC would lead the project, undertake all administration tasks, and oversee meetings and the timetable and progress of the various work packages.

The project had a collective budget of €442,591 and commenced in April 2015 with an initial 12 month timescale, though this was extended to 14 months in recognition of the existing workload on staff in scientific institutions. A similar call for proposals was addressed to MS from the Mediterranean and Black sea regions and links were encouraged between the two successful consortia.

Work Package 1

Work Package 1 embraced three deliverables; element 1 (analysis of main coordination activities), 5 (regional work programme) and 6 (consultation process) in the call for proposals. Results for the three deliverables are briefly summarized below. Full texts are found in Annexes 1, 2 and 3.

WP1.1 Analysis of main coordination activities and tools for fisheries sampling plans achievements and areas of improvement

The countries in the North Sea region have a long tradition, within the ICES framework, to coordinate and cooperate regionally for the purpose of assessment of fish stocks. Fishery dependent data supporting assessments are however presently collected through national sampling programmes.

A large number of activities have been conducted to improve regional cooperation and coordination since the onset of the first EC data collection regulation in 2002. The most prominent one is the establishment of the regional coordination meetings with the specific objectives to identify areas for standardisation and to support collaboration and cooperation between Member States (MS). Substantial work, in particular to improve quality procedures, has also been carried out within the ICES framework. Important activities include the ICES Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS), workshops initiated by the group and successive groups (ICES WGCATCH, WGBIOP and PGDATA). Regional coordination and cooperation have been further supported by the development of certain tools. The most important ones are the Regional Database (RDB), the COST (Common Open Source Tool for raising and estimating properties of statistical estimates) tools and the WebGR (Web services for support of Growth and Reproduction studies). The latter two were funded by the Commission as studies.

Major achievements by regional coordination activities are:

- A common understanding of the fishery activities on a regional scale
- Common nomenclature for métier /fisheries, ports, species and areas and agreed reference lists
- Implementation of the regional database
- Trust building between MS
- Increased awareness of the need for statistically sound sampling schemes. Body of knowledge developed.
- Establishment of networks to improve quality assurance in age and maturity and maturity readings within and between MS.
- Improved efficiency in sampling activities through setting up bilateral agreements in the region.

Despite the steady progress in the field of cooperation between MS there is still room to improve cooperation between MS for the actual sampling of commercial fisheries. The reasons for this are discussed and include:

- An awareness of the need for statistically sound sampling scheme; From a regional cooperation perspective does it imply that regional cooperation for actual sampling moves from the concept of simple task sharing to integrated regional sampling designs? The latter require a much stronger theoretical background and integrated approaches to protocols, quality assurance procedures and data management.

- Absence of clear road map on how to move towards regional sampling plans
- Absence of long term funding for maintenance and further development of tools needed at the regional level
- Lack of funding for development of the regional database

The full report relating to the review of the regional coordination activities is available in Annex 1.

WP1.2 Towards regional sampling plans – a regional work programme

A regional sampling plan includes: agreed objectives based on end-user needs, an integrated regional sampling design, standardized sampling protocols, a common approach to quality assurance, and regional tools for the management and dissemination of data. All these different aspects need to be developed before a regional sampling plan can be implemented, and the fishPi project (WP2-4) has significantly contributed to several of these elements.

It is more than likely that a regional plan will be constantly tuned after its implementation as fisheries and landing patterns change and science evolves. Management of the plan and its implementation are expected to be a task for the future Regional Coordination Groups (RCG).

In the work programme we describe the different aspects and elements of these foreseen regional sampling plans, including the fishPi contribution. Short term (2016-2017) development needs, recognized by other WPs in fishPi or elsewhere, are identified. After the termination of fishPi there are no regional resources available for further development of such plans. Heads of Institutes and National Correspondents (NCs) have in the fishPi consultation questionnaire (Annex 3) shown willingness to contribute with expertise and staff time, but in several cases lack resources to do so. Prioritization needs to be completed by the relevant RCMs/RCGs.

The full report relating to the regional work programme is available in Annex 2.

Short term needs and processes for different aspects and elements of the development of a regional sampling plan are identified in the regional work programme and include (by topic):

Management and development of the regional sampling plans

- RCMs to evaluate the expertise and resources needed for the Regional Coordination Groups to develop regional sampling plans.
- RCMs to identify a process of collaboration with end users, and expert groups to help in development of elements of the sampling plans.
- Further work in a new project or under the umbrella of the RCGs to further refine regional sampling designs.

Regional designs of sampling plan for commercial fisheries (based on outcomes from WP2)

In the short-term, a scientific planning group for key fisheries should be identified and established to oversee the development of the regional sampling schemes. This will be a step-wise iterative approach, which will need to focus on sections of the fleet such as pelagic or demersal trawlers, and will need to consider the provision of data for the stocks fished by these fleets.

The results of the simulation studies for commercial fisheries in WP2 have indicated implications of regional sampling designs. However more research is required to improve the evidence for the move to regional sampling schemes, and to determine the appropriate effort allocation to meet agreed objectives. This will be a step-wise iterative process involving collaboration between all relevant parties.

The results of the simulation studies have highlighted several issues that are likely to be of concern to Member States and would need to be resolved between participating MS (and institutes) before a regional sampling design could be introduced. Discussions in RCGs are required to resolve these issues, which include:

- The regional allocation of effort would require redistribution of effort from one country (or fleet) to another.
- The sampling of landings from foreign vessels.
- The collection of biological data for species of less importance to the country.
- The mechanism for inclusion of national requirements in a regional sampling design.
- The appointment of groups that focus on the design & implementation of a certain fisheries sector in a long-term setting.

Regional sampling plans for estimating impact of fishing activities on the marine ecosystems, small scale fisheries and recreational fisheries (based on outcomes from WP3)

Before regional sampling plans can be designed and agreed in the RCGs a substantial amount of work is needed between the main end users, data collectors and experts in sampling, under the umbrella of the RCGs to identify:

- Priorities on data needs and criteria (i.e. set by-catch limits needed for designing monitoring)
- Standardize guidelines and protocols
- Monitoring methods taking into account cost/benefits analysis
- Practical implementation considerations
- High risk by-catch métiers by region in case of PETS
- For small scale fisheries and recreational fisheries evaluate national requirements for data to support national inshore management schemes, and find a trade-off between regional and national needs.

Regional databases and the management of data collected through regional sampling plans

- Further development of the Regional Database to meet requirements for future regional designs, statistically sound sampling (WP2) quality assurance checks (WP4), estimation procedures and dissemination.

Regional sampling plans and a Quality Assurance Framework (based on fishPi WP4)

The fishPi R package for quality checks are available in an openly accessible website (https://github.com/ldbk/fishPifct).

- A number of countries have volunteered to export their data into the fishPi data format and run the quality checks, in order to test the codes and report back to fishPi developers for improvement and potential new quality checks.
- WKRDB 2016 should trial the data format and data structure.
- All MS should then be encouraged to export their data into the fishPi format (Annex 9) and test the codes developed in fishPi. They should then report to RCGs for feedback on their conclusions and suggestions for improvements.
- Quality checks to be implemented in the RDB and tested on regional datasets.

Sufficient resources need to be allocated for development of the regional sampling plans. The consultation process within fishPi WP1 revealed that there is a substantial amount of expertise on different aspects of regional sampling plans available in different institutes and MS. There is willingness to support the regional work but resources in terms of staff-time and funding are often limited. It is important to realize that development of regional workplans is a step-wise iterative process involving consultation and collaboration with different parties ranging from data collectors, coordinators and main end-users. Robust resources to assure momentum are needed for several elements in the development of regional plans and need to be secured. Two key aspects are:

- EU to support a statistician team with time allocated to regional tasks.
- EU to support funding for further development of the Regional Database to meet requirements for future regional designs and estimation procedures.

WP1.3 Consultation on regional sampling plans

Member States have worked together since 2004 in Regional Coordination Meetings (RCMs) to coordinate national data collection programmes under the EU Data Collection Framework (DCF). In recent years it became clear that regional data collection plans would offer a number of benefits over non-coordinated national sampling plans. These benefits include more appropriate sampling designs, more rigorous statistical methodology, harmonised data formats and the adoption of standardised protocols for collecting, storing and analysing data. Additionally, more cost-effective data collection is a likely result from more efficient and better coordinated programmes. The establishment of regional sampling plans has, however, direct and indirect implications on the way data collection is currently organised between and within countries, and on the way data is stored, processed and disseminated to end-users. A key aim of the fishPi project was to support this transition.

An important part of the project was to undertake a regional consultation to compile views on possible future regional sampling plans as well as on different prerequisites, identified within fishPi, required to develop and implement such plans, determining the degree of consensus. Available expertise and resources required to develop and implement regional plans were further examined.

The consultation process was completed by written procedure through a questionnaire. The questionnaire was sent out to National Correspondents (NC) for data collection in MS and to heads of institutes from EU Member States fishing in ICES European waters (North Sea, North Atlantic and Baltic) as well as to ICES and the European Commission as they were considered the main end-users of regional data. Responses were recieved from National Correspondents or the institutes (in some cases both) in 15 out of 17 Member States.

The main results are summarized below. For full results and the questionnaire see Annex 3.

Overall, there was a high level of agreement in views expressed by NCs and heads of institutes. There was strong support and consensus among the responding institutes and NCs on the general concepts of regional sampling plans based on statistical designs. Respondents agreed with the general approach towards probability-based sampling plans and probabilistic selection methods. Regional sampling plans are considered the best approach to improve quality of data and the cost-effectiveness of the current national sampling plans.

However, not everything can be regionalized. There was a strong consensus on the need to maintain elements addressing specific national requirements or objectives within the national plans even if the overall design is regional. Therefore, it is important to make sure that regional and national data collection plans and objectives are complementary.

There is a substantial amount of expertise within key areas for development of regional plans, available at the institutes and MS. There is a willingness to make expertise available for development and implementation of regional plans but also some constraints, particularly where resources and available staff time are limited. Institutes and MS also have the expertise to run and interpret scripts developed in the statistical environment R, as the ones developed within WP4 of fishPi.

Most respondents foresaw different problems related to the sampling of landings made by foreign vessels in national harbours. The main problem seems to be robust access to fish. Reasons for problems with access to fish are both administrative (permits to sample the fish) and logistical (fish rapidly transported to trucks). A second problem is access for the sampling country to relevant data contained in logbooks and sales slips. A third problem is that sampling of landings requires effective regional data management for samples to be processed and raised properly. In summary, sampling of foreign vessels requires more attention and dedicated actions.

There a strong consensus and support from the institutes and MS on the need for further development and implementation of a multi-purpose regional database (RDB). Some respondents see the RDB as a prerequisite for regional sampling plans. Significant cost saving associated with having a functional regional database where end-users can access data and remove the need for specific data calls/requests are also highlighted. Some respondents stress that national expertise needs to be involved in the development and implementation of estimation procedures within the RDB. There is also a strong consensus on the need to develop the RDB further to include more quality assurance checks (as the ones developed within fishPi WP4).

Most respondents argue that additional resources will be needed to sample protected, endangered and threatened species (PETS) and/or stomach contents. The amount of additional resources is dependent on the level of ambition within the sampling plans. Priority levels need to be discussed and end-user needs have to be clearly defined. Respondents foresaw (or have experienced) some problems with access to vessels for PETS sampling. Foreseen problems are mostly related to potential refusals due to lack of space for observers on small vessels. Alternative sampling approaches might also be needed. All respondents answer that they have (or partly have, depending on the taxon) the relevant expertise for PETS identification. Some respondents highlight that training might be needed for observers. Expertise in stomach analysis is available in some institutes and MS. Resources to carry it out routinely and/or in large scale are however lacking.

Work Package 2

Work Package 2 was focussed on developing the regional sampling designs, and the formats and software that would underpin the implementation of those designs. The work package was divided into three sections, each with distinct but interrelated deliverables.

WP2.1 Sampling design principles, data set and estimation codes

This part of the work package was tasked with providing text that summarised the statistical principles that underpin the sampling design, the estimation and the sampling protocols. It was tasked with drafting a format suitable for the compilation of the regional data set to be used for simulation testing of regional designs and it was tasked with writing the R scripts and functions needed to provide estimates from the simulation models, to simulate the random and multi-stage sampling and provide the bias and precision indicators for these prospective designs.

The text “Principles in the Implementation of a probability based sampling design” is presented as Annex 4.

A data sharing agreement (Annex 5) was drawn up by MASTS and circulated to all project partners to be signed by the heads of the respective scientific institutions and administrations prior to the request for logbook and sales note data. This data sharing agreement was effective from 10th April 2015. The agreement specified the usage of the data and the protocols for storage; the latter including a password protected limited access area on the project SharePoint.

The format for which logbook and sales note data was requested was finalised in May 2015 and circulated in June. It is presented in Annex 6. Data from all 10 countries participating in the project was provided and uploaded to the project SharePoint in time for the June case study start-up meeting. During and subsequent to this meeting, a large amount of checking and analysis code was written and shared among project members of the case studies in order to refine and hone the data sets being used. Examples of this manipulation, checking and analysis code are presented in Annex 7.

Each case study compiled a data set specific to their needs of their simulation models, and in most instances specific simulation and estimation code written for each case study. These are presented as an annex of the case study reports. More generic code, developed and shared during the start-up meeting in June is presented in Annex 8.

WP2.2 Formats and code lists

This part of the work package was tasked with providing the R code to generate the data exchange format needed for design based data collection, compatible with the operation of the checking and QA functions to be developed in work package 4. Additionally it was tasked with providing a comprehensive set of harmonised code lists needed for the population of the data exchange format and the aggregation of regional data sets.

Work Package 2.2 produced R packages and code lists associated with the data exchange format for commercial sampling data. This format is needed for storing sampling data collected under stratified design based sampling schemes using multi stage probability based selection methods. This format is a development of the data exchange format used in the COST project (COST 2006), and for the

Regional Database (RDB). It was further developed and used during the WKRDB 2014-01 (ICES 2014) workshop in October 2014 in Aberdeen, UK, and the WKRDB 2015-01 workshop in Sète, France. For development and testing during the project the format was termed "csPi". It should be noted that this format will evolve as more nations adopt probability based selection methods and estimation routines. The core contributors to WP2.2 were also involved in the simulation modelling in WP 2.1, and the data checking WP4. Members of the Mediterranean and Black sea project were also involved in developing the format.

The "fishPiFormats" package contains an R object version of the format. The fishPiCodes package contains code lists for areas, species using the ASFIS codes, species using the WoRMS codes, métiers, UN locode location codes and vessel type codes. Both these packages are available from the project SharePoint http://www.ices.dk/sites/Projects/fishPi/Rpackage

A description of the format and the PDF version of the „fishPiformats" package is presented in Annex 9.

WP2.3 Regional fisheries sampling design case studies

This work package was tasked with running simulation models of the sampling to test possible regional sampling designs for the fisheries exploiting regional stocks. This was achieved through four case studies of commercial fisheries (pelagic, demersal, flatfish and northern and southern hake stocks). Each case study considered a number of regional sampling designs, as a first attempt at objectively selecting designs best suited to achieve regional goals, and highlighting some of the considerations that would be involved in their implementation. Sampling designs were compared by using simulation studies to mimic the random selection of fishing trips from the regional population of trips for that fishery.

The data sets used to achieve this, consisted of the aggregated logbook and sales note data provided by 10 countries involved in the project: Belgium, Denmark, England, Ireland, Scotland, Sweden, Netherlands, France, Spain, and Portugal. Collating these data sets proved to be a far more ambitious and complex task than was reflected in the original proposal. These data were at the level of the fishing trip, and enabled simulation code to be written to mimic trip level data collection: for on-shore designs this was a two-stage sampling of port day, and then vessel; for at-sea designs a single stage sampling of a trip by a particular vessel. Various stratified designs were devised based on the characteristics of the landing locations and the vessels. Some designs had a national component to their stratification, others not. For each simulated data set an estimate of total landed weights by various domains (species, country etc.) were obtained using the Horvitz Thompson estimator (Horvitz and Thompson, 1952) from the R package "survey" or independently derived R code. Repeated simulations enabled the performance of the designs over many iterations to be assessed, the mean estimates of the totals being compared to the known "true" value from the data set, and the variability of the estimates determined from the relative standard errors. The species landed weight was used as a proxy for the biological data (age and length distributions) that are actually collected. The design effect, relative to a simple random sample, was evaluated to compare the merits of the different designs. The total available effort was based on the number of on-shore sampling visits and at-sea observer trips undertaken at present by the countries and scientific institutions involved in sampling, with various effort allocation scenarios considering how this effort could be optimised.

Four recognisably distinct fisheries that operated on shared stocks were considered as suitable candidates for cooperative regional designs. These case studies were:

- Case study 1 Fisheries for small pelagic fish operating in the North Sea and North Atlantic
- Case study 2 Fisheries for demersal species in the northern North Sea
- Case study 3 Flatfish fisheries in the North Sea area (includes North Sea, Skagerrak and Eastern Channel)
- Case study 4 Fisheries for hake, both Northern and Southern stocks, in the North Atlantic

Each of these case studies drafted a brief description of the fishing fleets, the species and stocks, and the arrangements for the sampling and assessment of the main stocks at present.

The main conclusions relating to regional designs were as follows:

CS1 Pelagic fisheries

The main findings in the small pelagic case study were that with a relatively simple stratification and the present regional sampling effort it would be possible to achieve a very effective regional sampling design.

The small pelagic case study considered both port sampling designs and at-sea sampling designs. For the onshore sampling design one of the findings was that in terms of stratification the simulations indicated that using the vessel length as strata greatly improved the sampling design. This was because the bulk of the landings were accounted for by large landings from relatively few large vessels. However, to be able to implement this design, every sampling site would need a list of all vessels above 40 metres. Contrary to expectations, stratifying by large and small sampling sites did not improve the sampling design in comparison to when countries were used as strata, only when sampling site and country were both used as strata did it improve the design. The major finding of the on-shore design was that the present allocation of sampling effort across countries and within a country could be optimised if sampling effort could be freely redistributed across countries. Within a country nearly all countries could gain in the sampling design if the vessel length were to be considered when the sampling schemes were conducted. Furthermore, if a sampling country only sample vessels from their own flag country, then a large proportion of the landings will never be available for sampling; in 2014 only 59% of the landings from a vessel flag country were landed in the country of the vessel. Besides, in a regional design, EU landings in third countries needs to be taken into account as in this case study landings outside the EU summed up to 17% of the total landing.

For at-sea sampling there were several clear advantages to self-sampling, which if well designed and conducted could be a very cost effective way to sample, although it should be noted that control samples will also be needed to ensure independent data. As it is an industrial fishery there is the advantage that samples can be frozen directly after the catch and thus obtained on a haul by haul basis and not on a trip level that a harbour sample will most likely be. Self-samples were considered likely to be in a better condition than those from a port sample. A self-sampling design would sample the flag vessel as long as the vessel can keep the sample on board until it arrives in the home harbour. The effort allocation between the countries would need to be optimised and re-distributed to a more optimal coverage.

An assumption in this case study has been that the landings can actually be sampled unsorted at factories, but there is no clear indication of landing sites (factory/ marked/ fishing auction etc.) or degree of sorting in the data used in this project – and it may even be problematic to achieve this information on a national level. This information will need to be improved to achieve an optimal sampling strategy for small pelagic fisheries in the future.

The case study 1 report is presented in Annex 11.

CS2 Demersal fisheries

Of the alternative designs explored in the North Sea demersal case study, a country and major and minor port design performs best in the regional context. Major ports were defined as the 43 ports receiving 95% of the landed catch, while the remaining 321 ports made up the minor ports. All countries within the North Sea would have sampling commitments under such a regional design and it would potentially generate unbiased and more precise estimates for individual fish species, and national estimates than those obtained at present. The major implication of this design is that given the distribution of landings, there would be substantial reallocation of sampling effort between different countries. The feasibility of such would need to be considered both nationally and within the regional administration.

The simulations also suggested that the present sampling designs and effort allocation do not cover the full distribution of on-shore landing locations, to the extent that estimates for species are potentially biased due to lack of coverage. The effort allocated to national schemes at present is far from optimal, with some countries obtaining precise estimates due to high sampling levels, but at the expense of the regional estimates.

For at-sea scheme designs based on national fleets, stratification by vessel length, and a combination of national fleet and vessel length were compared. It was found that the national and vessel length stratification provided the best of these designs and that, as with the on-shore schemes, optimal effort allocation would require substantial changes in the existing situation.

The case study 2 report is presented in Annex 12.

CS3 Flatfish fisheries

The flatfish case study focused on the on-shore sampling, where market-visits and trips are the main sampling frames. Beam trawlers, demersal trawlers, and gillnetters are the dominant vessel types and account for 95% of the total flatfish catch. For the on-shore design it was found that the landing composition of a harbour usually contains vessels from multiple countries. A comparison of two sampling schemes, a national sampling scheme in which only flagged vessels for the country were sampled by that country, and a regional scheme, in which all landings into a country were sampled, showed that the regional sampling scheme gives unbiased estimates, while the estimates from the national scheme were biased. Therefore, a clear advantage of a regional sampling scheme would be to establish the practice of sampling all the international trips, rather than the trips belonging to the country of the landing harbour. The selected optimal sampling scheme in the case study was stratified by quarter and vessel length with sampling effort allocated according to the Neyman-allocation, and post-stratification by area. A transfer to a regional sampling scheme based on this scenario would imply a significant change in numbers of trips sampled per country, however the current national sampling programmes are not based on flat fish sampling only.
The case study 3 report is presented in Annex 13.

CS4 Hake Fisheries

The hake case study considered both the sampling of the southern stock, and the wider stock involving the northern components. It also considered designs that focused on the sampling of hake for the trips where hake was present (positive trips), and designs based on sampling of all trips regardless of the presence of hake (all trips). Simulations were carried out using simple random sampling with replacement. Year 2013 was used to derive the stratification and allocate effort proportional to landings, and 2014 was used to simulate and evaluate the regional scenarios. The general conclusion was that regional sampling designs improved the precision of the estimates in this fishery, with the most precise estimates of total catch being obtained in scenarios stratified by port and, secondly, by port and country and by port and quarter. The case study also considered the wider implications of achieving good coverage by country, and also by other domains (stocks, métiers, temporal periods), and the extent that these could be compromised if a regional design is simply derived from statistical optimization of landings estimates of a single species and from simulation analysis that include only positive trips. Feasibility analysis revealed the need to consider the extent to which particular countries with smaller contributions to overall landings of hake would see their sampling plans reduced as this would compromise other requirements for advice such as those related to other stocks or local management measures established by National governments.

The case study 4 report is presented in Annex 14.

WP2 Conclusions

The exercise undertaken in WP2 has highlighted the main stages that are needed in the development of regional sampling schemes for the collection of biological data for regional fisheries and the generation of valid statistical estimates from that data. In summary these involve the following main tasks:

- Development of regional statistically-based sampling designs and related estimation methods to achieve regional objectives.
- The agreement of those regional objectives with the end-users of the data.
- Agreement between the participating scientific institutions (and nation states) on the data to be collected, the regional collection protocols to be adopted, and the estimation methods to be used.
- The resolution of implementation issues such as the inclusion of national requirements, the sampling of foreign landings, the distribution of regional sampling effort.
- Effective and efficient data management - harmonised data formats, adoption of agreed code lists, dissemination of common software tools, a regional database and clear and efficient data flow channels from national to regional databases for sampling data collected by scientific institutions, and population data collected by national control agencies.

The development of regional statistically-based sampling designs can be most effectively carried out by regional scientific bodies, either along the lines of those established under the auspices of ICES for research vessel surveys (e.g. WGIBTS, WGIPS), or as regional scientific groupings operating directly under the auspices of the RCGs. These regional survey groups would focus on a species fishery or group of fisheries, for example North Sea demersal fisheries, North Atlantic demersal fisheries, North Atlantic pelagic fisheries, North Sea shellfish fisheries, small scale fisheries. These bodies would develop the science behind the surveys and their practical implementation, and could develop improvements to the survey design over time exactly as for the research vessel surveys. As this project has demonstrated, the process of assembling the scientific and statistical experts for regional scientific bodies needs to be an inclusive not exclusive process. There is also a role for an enhanced regional body tasked with development and dissemination of enabling software and facilitating national and regional data management. The link between the different planning groups, data management and between the science and the data collection regulation could then be provided by the Regional Coordination Groups.

Sampling schemes for mixed-species fisheries are unlikely to be able to simultaneously meet objectives for all species and some discussion regarding prioritising objectives will need to be carried out both with the EU and the scientific end-users of the data. Regulations regarding collection of fisheries-related data need to be flexible enough to allow the above developments, and to devolve the scientific development of the data collection methods to the regional scientific bodies responsible for that work, as is currently the case for research vessel surveys.

The work of WP2 has demonstrated that there is a desire among scientific institutions to work collaboratively to improve the scope and quality of the data collected for regional fisheries. The case studies have demonstrated that there are some clear statistical and methodological benefits to regional sampling designs, both in the improvement of estimates and in the better allocation of whatever resources are available for fisheries data collection. What is now required are the mechanisms to persuade MS to participate in regional sampling schemes. Most obviously these would be clear demonstrations of the tangible benefits of regional organisation of sampling to individual scientific institutions, the wider fisheries community and Member States.

Work Package 3

Protected, endangered and threatened species (PETS) and stomach content data collection is not well covered under the current Data Collection Framework (DCF). fishPi WP3 focused on how future regional programmes should be carried out to collect these variables. A regional sampling programme has been described and the work to be carried out in the different stages analysed. RCGs, experts in fisheries monitoring and end-user coordination and roles have also been discussed.

In the case of SSF/RFS, fishPi focused on the current state of knowledge of SSF and RFS in Europe, the drivers and end users of data from these fisheries, the sampling methods that can be used, and how national sampling programmes could be coordinated regionally. Information is drawn from previous studies and from reports of ICES expert groups dealing with data collection from commercial and recreational fisheries.

The main conclusions from WP3 are presented below.

WP3.1 Small Scale and Recreational Fisheries

An unambiguous definition of SSF would help with managing the sector and implementing targeted policies. The fishPi project agrees with the 2013 DCF (Nantes) workshop that SSF should refer to <10m (LOA) vessels, for which there is no Control Regulation requirement to submit EU logbooks, and to a separate 10-12m category which is excluded from the VMS regulation. fishPi also noted that many countries have exemptions from VMS regulations for some vessels of 12-15m LOA.

The types of data needed from SSF and RFS should be defined by end-user needs to achieve Common Fisheries Policy goals as well as to support national inshore management. Marine Spatial Planning and the Marine Strategy Framework Directive also define the types and resolution of data to be collected. Data on structure, activities and catches of the fisheries are also needed to help design effective and enforceable control measures and to monitor the outcomes.

Where census data are not available for SSF or RFS, surveys are needed to estimate catches, effort, or other information needed for stock assessment and fishery management. The surveys normally involve two stages: firstly, a survey to describe and quantify the total population of vessels or fishers, and secondly, selection of a representative sample of the population to collect more detailed data on catches, effort, gears used and other information needed. The survey design depends on what data are already available, such as the existence of vessel registers or lists of recreational license holders.

A diverse range of survey methods are possible and are described in this report. Essential requirements for each national survey are: i) a fully documented, statistically-sound survey design and analysis aligned with end user needs; ii) a quality assurance procedure to ensure that survey protocols are followed and that archived data are quality-controlled; iii) full documentation of issues arising during implementation, such as non-responses or incomplete coverage, to allow evaluation of potential for bias; and iv) provision of quality indicators needed by end users. These requirements are the same as for sampling schemes for commercial fisheries to estimate length or age structure or estimate discards, and described in fishPi Work Packages 2 and 4.

The range of survey designs for RFS is wider than for SSF, for example including nationwide population screening surveys, and the collection of catch data using off-site surveys (e.g. catch diaries) or on-site surveys (intercept and roving-creel surveys). The appropriate design of a sampling program for recreational or small scale fisheries in a region does not necessarily have to be completely harmonized between countries. The most important attribute is that surveys have robust statistical designs to minimise bias and allow correct calculation of precision. Catch estimates from different surveys can then be reliably combined.

WP3.2 Stomach contents

Some MS already collect diet information, but it is not a general practice and in the majority of the cases the sampling is not coordinated at a regional scale and the information obtained is not available for the scientific community.

There is also considerable historic data mostly from pelagic and demersal commercial species that could be integrated in common regional datasets to feed existing models and understanding long-term community interactions within each ecosystem.

Defining a sampling plan using a single stock has no added-value as it does not provide enough information to move towards the Ecosystem Approach to Fisheries (EAF) with sufficient, statistically-sound samples of all the key species needed for food web characterization. The most effective sampling scheme is a hostage to specific user needs and highly dependent on the species considered. These should be agreed between the MS scientific community based on some general principles given in this deliverable. Many of the sampling guidelines suggested highly benefit from an opportunistic sampling in internationally coordinated surveys inside the DCF and add on fish diet sampling to minimize direct costs, providing comprehensive and comparable diet description on a regional basis.

WP3.3 PETS

Nowadays, data availability on PETS by-catch data is quite scarce, not standardised and not included in a specific database that could be used for the design of regional sampling programmes. At present, the main sources of data available from PETS come from the current DCF and Council Regulation (EC) No.812/2004.

Although there is a wish to monitor a broad range of species, covering several taxa, an overarching design that adequately covers all taxa within the DCF is not realistic. When incorporating monitoring of PETS in the new DCF/ DC MAP, the emphasis should therefore be on improving on board sampling protocols to ensure PETS by-catches are captured within the data recording system and to alter downstream data handling systems to ensure by-catch records of PETS are easily accessed by end users.

One approach to help address some of these issues may be to use data collected under the DCF or other sources to help identify “hot spots”, such as areas, seasons or métiers with relatively high by-catch rates of PETS. Based on initial assessments of the data at this larger scale, relevant Member States or RCG’s may then need to carry out more focussed surveys to fully assess the scale and patterns of PETS by-catch in specific fisheries. This approach would require Member States or RCM’s to identify additional fisheries and/or species requiring sampling.

A combination of scientific observers at sea programmes and REM methodologies will probably be the best approach when directed PETS by-catch monitoring is needed as cost effective and optimized sampling programmes.

Future Needs

Regional coordination of the three programmes covered under this WP (PETS, stomach content and SSF/RFS) will improve the quality and cost-effectiveness of the data to be collected. Coordination should involve the lead scientists for the surveys in each country in liaison with the Regional Coordination Group and ICES Working Groups on fishery data collection. An objective procedure for advising on relative survey effort in each country would need to be developed and tested by simulation as done in fishPi WP2 for LSF sampling.

In the case of PETS, on-board sampling coverage and protocols should be adapted to cover those areas, seasons or métiers with relatively high by-catch rates of PETS.

In the case of stomach content data, synergies between a stomach collection protocol and the monitoring of human pressures and impacts under the Marine Strategy Framework Directive (MSFD descriptors, D1, D4 and D10), and surveillance of marine biotoxins should be considered. This valuable additional information could be analysed at very little additional cost but relevant in many areas of scientific knowledge and with significant added value for the fishing industry, economies and human health.

In the case of SSF/RF, this will require knowledge of what fraction of the total regional harvests of each stock is attributable to SSF and RFS in each country, considering catches of all species subject to assessment and management in a region. In most cases, surveys can collect multispecies catch data with relatively little extra effort, but regional RFS catch estimates are currently available for very few species.

As a first step to regional coordination, data collection programmes for the variables covered in this WP (PETS, Stomach content and SSF/RFS), substantial work needs to be carried out between the main end users, data collectors and experts in sampling, under the umbrella of the RCGs, to: i) identify priorities on data needs and criteria (i.e. for shared international stocks in the case of SSF/RF); ii) identify additional national requirements for data to support national inshore management schemes, and find a trade-off between regional and national needs; iii) identify appropriate data collection methods for each country taking into account cost/benefit analysis and practical implementation considerations; and iv) develop standardized guidelines and protocols. This work will lead into a second phase of development to design a regional data collection programme to cover needs identified in the first phase. This should be fully documented and submitted by the RCG for peer review by experts in survey design. The process outlined implies considerable intersessional work between end-users and data providers, led by RCGs, and for which some funding and resources will be needed. Expertise and training is needed on data collection and analysis methods and sharing of expertise and skills across countries.

The potential for a diverse range of survey methods for RFS in a region means that the concept of a regional database for recreational fishery data may not be easily achievable. A principal focus in the short term should be to ensure that data from national surveys are properly archived and quality assured and that the surveys are fully documented and transparent for each country contributing to a regionally coordinated recreational survey programme. This would be a pre-requisite to consideration of any regional database structures that could accommodate such diverse data.

In the case of PETS and stomach content data, common databases (RDB, DATRAS) should be the preferred selection to upload these data. A lot of work has been done during recent years to standardize format and protocols to upload this data into these databases.

The full report from Work Package 3 is presented in Annex 15.

Work Package 4

The objectives of WP4 were to develop guidelines to evaluate the quality of data at national and regional levels using shared tools and to agree on timetables for the implementation on annual quality checks. The work in this WP was conducted together with the WP4 of the same project run in the Mediterranean and Black Sea, as agreed at the kick-off meeting with the EU. The data structure (called csPi data structure) upon which the work was developed was an upgrade of the COST/FishFrame format (ICES, 2008), as agreed in the SC-RDB (ICES, 2014) and further developed in WP2. The source code of all the functions is available on a freely accessible website (https://github.com/ldbk/fishPifct), and it is possible to compile a library directly from the website.

The Guidelines on the use of the quality checks functions were structured as follows:

- Verification of the data structure. Import csv files into csPi structure with integrity control of the tables, and check every field against reference tables and agreed list of entries.
- Verification of the consistency of the information populated into the database, e.g. identification of trips without fishing operations, missing samples weight, …
- Advanced data checks. Smart outlier detection of outliers in the numerical fields, identification of errors in fishing operation positioning.
- Exploratory data analysis. Production of maps, figures and reports, adapted from the COST library exploratory analyses functions.

These functions and the library are to be considered as the first version of a library which aims to improve with time and serves the needs of any end-user at national or regional levels. Moreover, the data structure upon which the functions were developed is also likely to evolve in the near future, so the functions were developed in such a way to accommodate and allow regional coordination groups (RCG) the liberty to agree on any structure, field names, field boundaries or reference lists for any component of the data structure. With the same flexibility idea, the reporting procedure of the quality checks adds a column to each of the tables of the data structure and indicates the detected issues at the record level, when relevant.

A calendar for applying the Quality Control at a national and regional level was proposed, taking account of the end-user needs and the national constraints. The ideal time frame would require Member States to run the quality controls at their level at each quarter and validate all their datasets by the end of February. The first half of March would then be used to populate the regional database and run the quality checks at the regional level. Adaptation of this time frame will have to be considered, since field constraints make it impossible for all Member States to be in a position to transmit validated data for all of their datasets by the end of February.

The full report from Work Package 4 is presented in Annexes 18 and 19.

Deliverables

The major sections of work carried out in the course of completing the project are presented as annexes in this report. Some work were not specified in the proposal as deliverables, while in same cases specific deliverables are aggregated into a single annexes; the case studies being a case in point.

Table of deliverables			
WP1	ANNEX 1	Deliverable 1.1	Review of RCM work
WP1	ANNEX 2	Deliverable 1.2	Regional work plan
WP1	ANNEX 3	Deliverable 1.3	Consultation on regional sampling plans
WP2	ANNEX 4	Deliverable 2.1	Principles in the implementation of sampling designs.
WP2	ANNEX 5		Data sharing agreement
WP2	ANNEX 6	Deliverable 2.1	Case study data set format
WP2	ANNEX 7		Data compilation code
WP2	ANNEX 8	Deliverable 2.1	Simulation and estimation code
WP2	ANNEX 9	Deliverable 2.2	csPi format
WP2	ANNEX 10	Deliverable 2.2	code lists
WP2	ANNEX 11	Deliverable 2.3	CS1 Pelagic fisheries sampling designs
WP2	ANNEX 12	Deliverable 2.3	CS2 North Sea demersal fisheries sampling designs
WP2	ANNEX 13	Deliverable 2.3	CS3 North Sea flatfish sampling designs
WP2	ANNEX 14	Deliverable 2.3	CS4 Hake sampling designs
WP3	ANNEX 15	Deliverable 3.1	Sampling plan for the collection of PETS
WP3	ANNEX 16	Deliverable 3.2	Sampling plan for the collection of stomach contents data
WP3	ANNEX 17	Deliverable 3.3	Small scale fisheries & recreational fisheries designs
WP4	ANNEX 18	Deliverable 4.1, 4.4	Guidelines and calendar for quality indicator checks
WP4	ANNEX 19	Deliverable 4.2, 4.3	List of quality check functions

Dissemination Activities

Dissemination of fishPi project deliverables 2015-2016		
Presentation prepared for WGRFS	Sukarrieta	1st June 2015
Presentation to RCM NS&EA	Den Haag	1st September 2015
Presentation to RCM NA	Hamburg	13th September 2015
Presentation to Liaison meeting	Brussels	8th October 2015
Presentation to WK RDB 2015-01	Sète	26th 30th Oct 2015
Presentation of fishPi to WGCATCH	Lisbon	9th -13 November 2015
Presentation CS 1 to WGCATCH	Lisbon	9th -13 November 2015
Presentation CS 2 to WGCATCH	Lisbon	9th -13 November 2015
Presentation CS 3 to WGCATCH	Lisbon	9th -13 November 2015
Presentation CS 4 to WGCATCH	Lisbon	9th -13 November 2015
Presentation prepared for WGBYC	Copenhagen	1st February 2016
Presentation to PGDATA	San Sebastian	4th March 2016
CS 2 abstract submitted to ICES ASC	Riga	30th April 2016
CS 4 abstract submitted to ICES ASC	Riga	30th April 2016
WP 4 abstract submitted to ICES ASC	Riga	30th April 2016
WP3 abstract submitted to ICES ASC	Riga	30th April 2017
Presentation prepared for DCF expert group meeting	Brussels	2nd May 2016

Project Meetings

fishPi Meetings Achieved					
Number	**Meeting**	**Location**	**Date**	**Type**	**Attendees**
1	PMC meeting	on-line	10-Apr	GTM	MJ, ED, KR, JV, AP, LC
2	Kick Off	Brussels	20-Apr	FF	MJ, AP, JV, KR, EM, Commision, Med and BS
3	PMC meeting	Brussels	21-Apr	FF	AP, JV, KR, EM
4	Project start-up meeting	Aberdeen	5-6 May	FF	MJ, ED, AP, LC, KR, LZ,JV, JH, MC.
5	Statistical planning meeting	Aberdeen	07-May	FF	AP LC, JHV, MC, JV
6	Software development meeting	Aberdeen	08-May	FF	AP LC, JH, MC, JHV, KBH, (HKN, LD by GTM)
7	All person start-up	on-line	11-May	GTM	
8	CS leaders meeting	on-line	18-May	GTM	AP, LC, JC, MSP, EvH
9	WP2 start-up meeting	on-line	19th May	GTM	AP, LC, EvH, HG, HKN,JU, JW, JV, JE, JC,JR, JEg,
10	WP2 CS leaders meeting	Aberdeen	9-10 June	FF	AP,LC,JC, MSP, CC,
11	WP2 CS working meeting	Aberdeen	23-25 June	FF	AP,LC,LD,MC,KBH, LZ,JE,ARS, JC,PB,CC,NP,JR,JW
14	WP 2.3 simulation meeting	Aberdeen	26-Jun	FF	LC, MC
15	WP4 Meeting	Port-en-Bessin	6-8 July	FF	JV,AP,CC,GT(Med&B), MH,EV,KM,LD,ARS, MA,
16	WP3 Working Group Meeting	Sukarrieta	21-Jul	FF	EM, LZ,EA,ML, MA,AM, Simon Northridge WGBYC by GTM
17	fishPi & Med Black Sea liaison meeting	on-line	28-Jul	GTM	PMC of each bid
18	WP2.3 CS leader meeting	on-line	13-Aug	GTM	AP, LC, EvH, MSP, JR,PB
19	CS leders meeting	Aberdeen	25-Aug	FF	AP, LC
20	PMC meeting	on-line	22-Sep	GTM	AP, LC ,ED, EM, KR, JV
21	WP2.3 Progress core group meeting	on-line	22-Sep	GTM	AP, LC ,JC, EvH, CC, KBH, NP, MAz
22	CS1 CS2 combined progress meeting	Aberdeen	28 Sept 2 Oct	FF	AP, LC, MSP, KBH, CC, ARS, JE
23	WP1 Internal meeting	Brussels	7 Oct	FF	KR, JD, SV, JR, CS
24	CS3 case study meeting	IJmuiden	14-16 Oct	FF	EvH, CC, ET, JE, JW
25	CS4 case study meeting	Lisbon	14-16 Oct	FF	JC, JR, LZ, MA, LD, JP, NP (skype)
26	Mid term meeting	Brussels	20-21 Oct	FF	MJ, AP, EM, KR, COMM, Med & B
27	PMC and CS leaders meeting	Aberdeen	18 20 Nov	FF	JV, EM, LC, AP, JC, EvH, MSP,MJ, ED
28	PMC meeting	on-line	06-Nov	GTM	MJ, ED, AP, LC, EM,KR, JV
29	WP1 Internal meeting	on-line	30-Nov	GTM	MJ, ED, AP, LC, SV, CS, JD, JR
30	WP3 Working group meeting	Sukarrieta	19 Jan	FF	LZ, EA, HM, AM, PB, EM
31	WP1 Internal meeting	Gothenburg	15-18 Feb	FF	KR, LC, JV, JD, SV, ET, CS, MA, JR, MH, (EM, MJ, ED by GTM)
32	WP2 Internal meeting	on-line	23-Mar	GTM	LC, AP, MC
33	WP1 PMC liason meeting	on-line	01-Mar	GTM	KR, AP, LC
34	WP3 Internal meeting	on-line	27-Apr	GTM	LZ, EA, HM, AM, PB, EM
35	WP4 Internal meeting	on-line	28-Apr	GTM	JV, LD, AP, LC
36	WP1 Consultation process meeting	on-line	29-Apr	GTM	KR, ED, LC, JV, JD, SV, JR, ET
37	PMC Wrap-up meeting	Aberdeen	3-4 May	FF + GTM	MJ, ED, KR, JV, EM, AP, LC, (by GTM JC, MSP, EvH)
38	PMC Report coordination meeting	on-line	26-May	GTM	MJ, AP, LC, KR, JV, EM

*GTM – OnlineMeeting

Project Participants

Name	Institution	e-mail	Contribution
Mark James	MASTS	maj8@st-andrews.ac.uk	PMC
Emma Defew	MASTS	ecd2@st-andrews.ac.uk	PMC
Alastair Pout	MSS	a.pout@marlab.ac.uk	PMC, WP2, WP4
Liz Clarke	MSS	l.clarke@marlab.ac.uk	PMC, WP2, WP1, WP4
Katja Ringdahl	SLU	katja.ringdahl@slu.se	PMC, WP1
Estantis Mugerza	AZTI	emugerza@azti.es	PMC, WP3, WP1
Joel Vigneau	IFREMER	Joel.Vigneau@ifremer.fr	PMC, WP4, WP1
Christoph Stransky	TI-SF	christoph.stransky@thuenen.de	WP1
Jørgen Dalskov	DTU Aqua	jd@aqua.dtu.dk	WP1
Sieto Verver	IMARES	sieto.verver@wur.nl	WP1, WP2
Els Torreele	ILVO	els.torreele@ilvo.vlaanderen.be	WP1, WP2
Jose Rodriguez	IEO	jose.rodriguez@st.ieo.es	WP1, WP2
Maria Hansson	SLU	maria.hansson@fiskeriverket.se	WP1, WP4
Ana Ribeiro Santos	CEFAS	ana.ribeirosantos@cefas.co.uk	WP2, WP4
Hans Gerritsen	MI	hans.Gerritsen@Marine.ie	WP2
Henrik Kjems-Nielsen	ICES	henrikkn@ices.dk	WP2
Jens Ulleweit	TI-SF	jens.ulleweit@thuenen.de	WP2
Jon Elson	CEFAS	jon.elson@cefas.co.uk	WP2
Jon Helge Vølstad	IMR	jon.helge.voelstad@imr.no	WP2
José Castro Pampillón	IEO	jose.castro@vi.ieo.es	WP2
Julia Wischnewski	TI-SF	julia.wischnewski@thuenen.de	WP2
Kirsten Birch Håkansson	DTU Aqua	kih@aqua.dtu.dk	WP2
Manuela Azevedo	IPMA	mazevedo@ipma.pt	WP2
Marie Storr-Paulsen	DTU Aqua	msp@aqua.dtu.dk	WP2
Mary Christman	MCC	marycchristman@gmail.com	WP2
Nuno Prista	IPMA	nmprista@ipma.pt	WP2
Sofie Nimmegeers	ILVO	sofie.nimmegeers@ilvo.vlaanderen.	WP2
Edwin van Helmond	IMARES	edwin.vanhelmond@wur.nl	WP2
Lucia Zarauz	AZTI	lzarauz@azti.es	WP2, WP3
Patrik Börjesson	SLU	patrik.borjesson@slu.se	WP2, WP3
Chun Chen	IMARES	chun.chen@wur.nl	WP2, WP4
Ana Moreno Marques	IPMA	amoreno@ipma.pt	WP3
Eider Andonegi	AZTI	eandonegi@azti.es	WP3
Hugo Mendes	IPMA	hmendes@impa.pt	WP3
Simon Northridge	USTAN	spn1@st-andrews.ac.uk	WP3
Mike Armstrong	CEFAS	mike.armstrong@cefas.co.uk	WP3, WP1, WP4
Laurent Dubroca	IFREMER	laurent.dubroca@ifremer.fr	WP4, WP2
Norbert Billet	IRD	norbert.billet@ird.fr	WP4
Eva Velasco	IEO	eva.velasco@gi.ieo.es	WP4
Kelig Mahe	IFREMER	Kelig.Mahe@ifremer.fr	WP4
George Tserpes	HCMR	gtserpes@hcmr.gr	WP4

Photo Gallery of Selected Meetings

References

COST. 2006. *Studies and Pilot Projects for Carrying Out the Common Fisheries Policy Call for Proposal Ref : FISH/2006/15 – Lot 2 Project No :SI2.467814 Project Acronym :COST*. EUROPEAN COMMISSION.

Horvitz, D. G.; Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *Journal of the American Statistical Association*, 47, 663–685.

ICES 2014. Report of the workshop on developing the RDB data format for design based sampling and estimation (WKRDB 2014-01), 27-31 October 2014, Aberdeen. ICES CM 2014/ACOM:68

ANNEX 1 WP1 Analysis of coordination activities and tools for fisheries sampling plans

fishPi project (MARE/2014/19)

“STRENGTHENING REGIONAL COOPERATION IN FISHERIES DATA COLLECTION”

Deliverable 1.1 – Analysis of main coordination activities and tools for fisheries sampling plans – achievements and areas of improvement

Lead: SLU, Sweden.

Core Team: Katja Ringdahl, Maria Hansson (SLU, Sweden)
Mike Armstrong (CEFAS, UK England)
Jose Rodriguez (IEO, Spain)
Jörgen Dalskov (DTU-AQUA, Denmark)
Sieto Verver (IMARES, Netherlands)
Christoph Stransky (TI-SF, Germany)

Table of Contents

1 Introduction

The countries in the North Sea region have a long tradition, within the ICES framework, to coordinate and cooperate regionally for the purpose of assessment of fish stocks. Fishery independent data are also usually collected through surveys that are coordinated through planning groups such as ICES IBTSWG.

Fishery dependent data are however presently collected through national sampling programmes. This implies that it is a national responsibility to design and implement the sampling program, to introduce and maintain quality assurance procedures and to manage primary, detailed and aggregated data and estimates. The final national estimates that are delivered to end-users are in most cases point estimates. This means that quality in reality might be largely unknown for end-users and for other MS involved in the same fisheries or utilizing the same fish stock. Regional coordination and cooperation in sampling of fisheries can therefore be seen as an exercise to increase the overall knowledge and awareness of different sampling, analysis and fishing activities within a region with the overall objective to facilitate coherent sampling procedures and transparent quality evaluation. The final goal would be integrated regional designs and sampling plans, such as the ones simulated within the case studies of fishPi WP2. Regional sampling plans are expected to be cost-effective and increase the overall quality of data (fishPi questionnaire).

A large number of activities have been conducted to improve regional cooperation and coordination since the onset of the first EC data collection regulation in 2002. The most prominent one is the establishment of the regional coordination meetings with the specific objectives to identify areas for standardisation and to support collaboration and cooperation between MS. Substantial further work, in particular to improve quality procedures, has been carried out within the ICES framework, and include important activities such as the ICES Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS), workshops initiated by the group and successive groups (ICES WGCATCH, WGBIOP and PGDATA). Regional coordination and cooperation has further been supported by the development of certain tools. The most important ones are the regional database, the COST (Common Open Source Tool for raising and estimating properties of statistical estimates) tools and the WebGR (Web services for support of Growth and Reproduction studies). The latter two were funded by the Commission as studies.

The present document include a description of the evolution of work within the RCM NS&EA 2004-2015, main contributions of ICES from ICES expert groups as well as important tools. We then highlight achievements and identify areas where improvement can be made.

2 Regional coordination meetings

2.1 RCM - Background

The former EU Data Collection Regulation (EC 1543/2000 and the implementing regulation 1639/2001) established a framework for the collection of fishing capacity, effort, landings and discards, biological and economic data by Member States. It was intended that this programme would provide the basic data needed to evaluate the state of fishery resources and the fisheries sector. Although the Regulations set out the detailed requirements of the programme, and specifically require MS to demonstrate cooperation and task sharing (Article 3 (d)), no mechanism was specified to achieve this. Since ICES is one of the main users of the data collected by MS and as a consequence of the MoU between EC-ICES, ICES set up the Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS) to look at ways of ensuring regional coordination and cooperation in the collection of biological data. The PGCCDBS recognised that there was a need to develop regional groups where issues relating to the fisheries and data collection within the region could be more effectively discussed and cooperation developed. At its 2004 meeting, PGCCDBS recommended the formation of regional planning groups. The issue of regional coordination of the Data Collection Regulation (DCR) was also discussed at the meeting of the Committee for Fisheries and Aquaculture (CFA) (Brussels 21/10/2003). There was agreement that regional coordination would greatly increase the efficiency, effectiveness and integration of the various DCR National Programmes. The details of both national and international coordination were outlined in a working paper from the Commission. In this paper, the Commission outlined four original main areas for regional coordination (Baltic Sea, North Sea, Atlantic shelf and the Mediterranean waters). These areas broadly fit in with the Regional Advisory Councils (RAC's) which were in the process of being set up. Since the introduction of the RCMs, a RCM for long distance fisheries has also been established. The Commission suggested that the Regional Coordination Meetings (RCMs) should be held annually and involve National Correspondents and both biologists and economists from each MS involved in the DCR programme. The key objectives of the RCMs were to identify areas for standardisation, collaboration and cooperation between MS. The Commission also outlined how the RCMs would be expected to collaborate with other groups such as ICES and STECF. A final liaison meeting between chairs of different meetings dealing with DCR and later DCF was set up to analyse the final reports in order to guarantee good overall coordination. On the basis of the reports, this group would make recommendations to the Commission and MS. The DCR was revised in 2008 and reformed into the Data Collection Framework (DCF) (EC 199/2008, EC 665/2008 and 2008/949/EU (later revised to 2010/93/EU)). In these regulations and decisions the role of the RCMs and their tasks in regional coordination are clearly defined. The RCMs are assumed to get an even stronger mandate in the recast of the data collection regulation 199/2008 and will at the same time turn from meetings to groups, with the possibility to work throughout the year.

The obligation for Member States to cooperate with each other within the field of data collection is expressed in the new Common Fisheries Policy EC 1380/2013 article 25.5

Currently, financial support for participation in the RCMs is eligible under the European Maritime and Fisheries Fund (EC 508/2014) article 77.2(e). There is also a possibility for the Commission to support cooperation between MS under EMFF through direct funding (article 86.2 (f)). Previously, the RCMs were supported through the DCF.

2.2 Evolution of RCM NS&EA 2004-2015

The first RCM for the North Sea area was held in 2004. Since then one meeting has been convened every year. The North Sea area was originally defined as IIIaN, IV and VIId. The scope was expanded in 2005, to include division IIIaS (Kattegat) as well as ICES sub-areas I and II. The name was then changed to RCM North Sea and Eastern Arctic. From 2009 and onwards ICES sub-areas XII, XIV, division Va and the NAFO areas have been included in this RCM as well.

The key objectives of the RCMs were to identify areas for standardisation, collaboration and cooperation between MS. The early years of the RCM NS&EA were spent on identifying such areas. The meeting spent time on giving guidance on the set-up of bi- (or multi-) lateral agreements, in particular for sampling of landings for trans-shipment or first sale in foreign countries. Other areas of importance were the coordination of sampling programmes for growth, sexual maturity and fecundity. The RCM had concerns on sampling intensity, sampling outside the spawning season and harmonization of maturity determinations. To assess the effects of spatial and temporal deficiencies in the maturity sampling, RCM North Sea recommended setting up species-specific workshops to give guidance in collecting and analysing maturity data. These workshops were later dealt with within the framework of ICES PGCCDBS.

Time was further spent on compiling information on sampling intensities within the region as well as examining if the main fisheries were sampled for discards. A main problem was that there was no common nomenclature to name and describe fisheries, implying that regional overviews were difficult to make. In 2005 the RCM NS&EA also realized that a common database was a prerequisite to achieve cooperation between Member States and effective communication with end-users.

"*The RCM North Sea considers that progress on improving the effectiveness of sampling, integration of data between countries and provision of high quality information to the ICES assessment Working Groups is heavily dependent on the use of a common database."*

The RCM NS&EA 2005 realized that effective communication with the end-users was dependent on active data providers and suggested "*that the data contributors to the WG should provide standard reports and analysis, agreed with the WG, to enable easy evaluation of the data being used"*.

A main concern for the MS at the time was the quality requirements in the DCR stating that MS should meet different levels of precision in their sampling programmes. MS had difficulties to calculate precision correctly (as fishery sampling programmes usually are multi-stage and involve transformation of data through eg ALKs) but also to meet the precision levels required. The RCM NS&EA 2005 saw the need for standardization and did "*strongly supports the proposal for an expert group to begin the task of developing a set of tools within an open source framework which will provide analysis and precision calculations. These tools will be developed to attach to FishFrame and/or become part of an integrated system".* The expert group later became the COST project that developed standardized tools for exploration of data as well as for basic calculation of precision.

In 2006 preparatory work for the DCR revision was carried out. A large part of this work was the construction of "the fleet activity matrix" (later appendix IV in 2010/93/EU). The matrix was developed during a set of *ad-hoc* meetings in Nantes, France. The matrix was a comprehensive way to overview all the fishing activities in the European fleet and it gave MS a common nomenclature to describe fisheries allowing for regional overviews of fisheries and comparison of sampling coverage and intensities between countries. It was a prerequisite for the transition from a stock based approach in the DCR to a fishery based approach in DCF.

The RCM NS&EA 2006 acknowledged the work done at the *ad-hoc* Nantes meeting but had already realized that the large amount of cells in the matrix was not ideal for stratification of sampling programmes and could cause some unnecessary over-stratification on a regional scale. It was a further concern that the different cells in the matrix were defined differently in different MS which would jeopardize a regional sampling scheme built on the basis of cells in the matrix (regional métier based sampling). The RCM further discussed how the Nantes matrix would be translated into national sampling programmes. There were two main issues raised. Firstly, how many strata have to be sampled and units (effort, landings, value) to be chosen to decide which strata to be sampled. On a regional scale, the idea was that sampling could then be allocated to the different countries involved in the different fishing activities. The second issue that was discussed was how to translate "the proportion" of the activity in the different fishing activity strata into number of samples. Concerning this issue, the RCM relied on the future outcome of the COST project since target of precision was considered central when translating proportions to sampling intensities in the different strata.

In 2007 the RCM NS&EA continued to work on the applicability of the fishing activity matrix as a basis for a regional design. The large number of cells to be sampled as well as the fact that the initial analyses of the data showed problems with consistent allocation between MS to the different fishing activities raised concerns. RCM NS&EA realized the necessity of having data available before and during the meeting to ensure discussions on regional design of data collection were fruitful. The meeting also realized that ages for a certain amount of stocks were oversampled compared to targets in the DCF and suggested that age reading may be a field for task sharing to a larger extent.

In 2008 the present DCF (199/2008) came into force and the RCMs got a clearer mandate. The RCM NS&EA worked on the métier-related variables in contrasting the sampling programmes for landings and for discards with the international statistics on effort, landings and total value. It was the first complete regional overview of the fishing activities in the area. This comparison was meant to assess the completeness of sampling coverage for the métiers accounting for 90% of the fishing activities by fishing grounds. The evaluation of the data focused on the following points i) métiers not covered by sampling programme were pointed out and ii) discussions on compatibility of the fishing patterns between MS and within métiers as a basis for future task sharing. The RCM NS&EA was comparing the planned sampling and the fisheries statistics given in the NP proposals 2009-2010 on a regional level. The production of working tables took a large part of the meeting due to inconsistencies between MS in the naming of métiers / fishing grounds as well as units for the actual data.

During 2009 and 2010 the RCM NS&EA continued to produce regional rankings of métiers and fishing ground as well as contrasting the sampling programmes for landings and discards with the international statistics on effort, landings and total value. Where appropriate the RCM suggested actions to be taken at given métiers for different MS. The work was time consuming as the information was taken from the NPs 2009-2010 and required a lot of “tiding up” of names and units. Large effort was put on the MS to respect “naming conventions” and it was agreed that international datasets needed to be compiled prior to the meeting to utilize the meeting time effectively. The RCM were also concerned that there were national differences within métiers and realised that national descriptions (homogeneity of the métier, target species, spatial and temporal distribution etc.) of the ranked métiers were required in order to allow the RCM to evaluate the compatibility of the fishing activities and consequently possibilities for task-sharing.

RCM NS&EA 2010 also updated an international overview of sampling plans for ages sampling and reading, compiled from the NP proposals 2011-2013. It was agreed that there were several instances where a large number of ages were taken (compared to targets in DCF) but until such time as there are tools in place to provide detailed analyses on optimum sampling levels, both spatially and temporally, it would not be advisable to embark on reducing these numbers. The meeting compiled a list indicating which species were suitable for task sharing in relation to age determination and also an indication of the countries which had appropriate expertise. It was agreed that individual MS should then review the list of species they were ageing and, where required, should contact one of the nominated countries to consider the possibility of a bilateral agreement to ensure the ageing work could be completed.

Time was also spent on compiling information on sampling programmes for recreational fisheries in the different MS.

MS had in the past expressed a need to establish a regional data base to facilitate coordination and harmonization of data collection. Also, SGRN stressed the need for a regional data base for this purpose (Anon. 2009c). The harmonization would lead to a more efficient data collection and also allow monitoring of data quality. During the 2010 meeting the RCM NS&EA unanimously agreed that FishFrame was the obvious candidate for the RDB. The RCM agreed that a steering group needed to be set up together with RCM Baltic and RCM NA to start to realize the RDB.

During 2011 and 2012 most MS gradually uploaded more and more data into the regional database prior to RCM NS&EA. This resulted in more efficient use of meeting time but also larger requirements on harmonization in coding of species, harbours, métiers and areas. The RCM NS&EA provided suggestions for development of the exchange format to the RDB-SC but development was slow due to lack of funds.

The 2009 and 2010 workshops (WKPRECISE and WKMERGE) held within the ICES PGCCDBS framework concluded that the level-6 fleet métiers identified in Appendix IV of Decision 2010/93/EC were in many cases too dynamic and unpredictable to constitute sampling strata in which all the primary sampling units (trips) were identifiable in advance for allocating sampling effort. The workshops suggested métiers to be considered as domains, instead of strata, allowing for sufficient identification of sampling frames. This was the onset of the entire discussion around “statistical sound sampling schemes” and had large implications on the design of regional data collection programmes. RCM NS&EA were thereby awaiting more guidance from ICES methodological groups (ICES WKPICS 1-3) before planning regional sampling.

During 2012 RCM NS&EA expressed its view on the future data collection in the “Oostende declaration”. The vision was that “*End users will receive relevant, high quality data collected through an efficient regional basis*”. And the mission “*Data collectors will use statistically sound sampling schemes and operate under the guidance of Regional Coordination Groups, in which end-user priorities are agreed and the coordination of data collection takes place to meet those priorities. Commercial Fisheries Sampling*”.

By 2013 most sampling, effort and landing data were available through the RDB to the RCM NS&EA. That freed up meeting time to more rigorous checking and discussion on data quality, for example by comparing information in Eurostat, InterCatch and the RDB but also for an exercise on regional sampling design taking sound statistical principles into account. The exercise was an embryo for a regional sampling design for on-shore sampling using data from the RDB. The design stratified by groupings of species and stocks with a gross optimization based on the landed weight for 2012. It was envisaged that sampling of the ports identified in the regional list would be carried out by the observers operating in the nations in which the ports are situated. This exercise was to a certain degree the basis for the present fishPi project.

The RCM NS&EA 2013 also initiated a road map to achieve an efficient way to evolve towards requirements anticipated from the new upcoming data collection legislation. The initial road map was further developed by the RCM NA. The RCM NS&EA 2014 did however conclude that the actual implementation of the road map was hampered by absence of the new legislation, the lack of development of the RDB and the lack of establishment of the RCG process.

The RCM NS&EA 2014 further discussed implications of the landing obligation on sampling plans, observer schemes and concern of overall loss of quality in catch data.

Concern was also expressed for the deteriorating quality of catch data when the landing obligation is introduced. The meeting also spent considerable time identifying the processes that needed to be established for obtaining and demonstrating high quality data.

The landing obligation continued to raise concerns during RCM NS&EA 2015. Experience of the implementation in the Baltic suggested that, in some situations the BMS (below minimum landing size) fraction was being grossly under recorded in logbooks and/or was simply not available where the landing data were derived from sales notes and BMS fraction was not sold.

The RCM NS&EA was provided with data from all MS allowing for a proper analysis of regional fisheries. Data analysis was carried out, including some basic audits of the data within the RDB. This was supplemented by descriptions of fisheries within the region: NAFO area, Eastern Arctic area, northern North Sea demersal fisheries, southern North Sea flatfish and pelagic fisheries. Each of these descriptions included the identification of sampling frames of major landing harbours, the main national fleets by métier, the ranking of species tonnages, and maps of the fishing locations and landing ports.

Analysis of the landings abroad, and the extent to which the RDB held sampling data from flag vessels other than the landing country, showed that a considerable proportion of the landings were either not being sampled or the samples of this fraction could not be uploaded to the RDB. An analysis of the age data from the RDB was able to demonstrate the scope and the number of determined ages by species and country, and relate this to the proportion of the landings of the species concerned. While the number of age readings need not be directly related to the proportion of the landed catch, the findings were of interest in demonstrating potential for task sharing in age reading.

3 ICES and the PGCCDBS framework

A substantial investigation into the quality of fisheries sampling programmes, data and associated analysis has been conducted by the ICES Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS: ICES 2002 - 2014), in their role to promote the ICES Quality Assurance Framework for biological sampling (Nedreaas *et al* 2009), and by workshops and study groups established by PGCCDBS. The PGCCDBS was established in 2002 in response to the EC-ICES Memorandum of Understanding (MoU) requesting ICES to provide support for the EU Data Collection Framework (DCF; EC Reg. 199/2008, 665/2008; Decisions 2008/949/EC and 2010/93/EU). In addition to establishing protocols and standards for fish ageing and maturity determination, the PGCCDBS and its workshops and study groups have covered topics such as sampling and estimation for maturity ogives (WKMAT: ICES 2007a; WKMOG: ICES 2008b), accuracy of sampling data (WKPRECISE: ICES 2009a; WKACCU: ICES 2008a), discard raising procedures (WKDRP: ICES 2007b); design of commercial fishery sampling schemes (WKMERGE: ICES 2010; WKPICS: ICES 2011a, 2012b; 2013a; SGPIDS: ICES 2011b, 2012a, 2013b; WGCATCH: ICES 2014, 2015) and recreational fishery surveys (WKSMRF: ICES 2009b; WGRFS: ICES 2012-2015).

These ICES initiatives have had a progressive impact since the late 2000s in increasing the awareness within the ICES community of the need for statistically-sound sampling design rather than *ad-hoc* methods, and have developed an important and well-documented body of knowledge on fishery sampling design, implementation and analysis. An important component of this has been the development of guidelines for best practice as well as proposals for ways in which the quality of sampling programmes and the data gathered from them can be documented for a range of end users such as stock assessment scientists, regional coordination groups and the European Commission. The practical implementation of sampling schemes has been a major focus, particularly by WKPICS, SGPIDS, WGRFS and WGCATCH.

4 Tools contributing to regional coordination, cooperation and quality assurance

4.1 Regional Database (RDB)

The Regional database (RDB) was originally developed by DTU-Aqua, Denmark, but since 2012 has been hosted by ICES. The content of the database is governed by a steering committee with representatives from different RCMs and the host. Maintenance of the RDB is covered by the MoU between ICES and Commission but there is no funding available from the EU for development to meet needs arising from establishment of regional sampling plans based on statistically sound designs. ICES have supported some development needs. The database holds anonymized detailed data from sampling events as well as effort and landing data aggregated at a low level. The geographical scope is presently areas covered by RCM Baltic, RCM NS&EA and RCM NA. It is populated through yearly data calls from the RCM chairs. The utilization of the RDB has made the work within the RCMs far more effective. RCMs thereby consider the RDB a prerequisite for efficient work implementing systems and procedures to ease the coordination and cooperation of data collection on the regional scale. The RDB has however considerable potential to support Member States and RCMs in evaluating the quality of data and the generation of estimates from biological data needed by end users. This would allow for efficient, consistent and transparent responses to data calls and other types of reporting. RCM NS&EA have repeatedly (2012, 2013, 2014 and 2015) put forward study proposals with the aim to develop the RDB. These study proposals were never granted.

4.2 COST

COST (Common Open Source Tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation) was a project funded through an EU study (FISH/2006/15 lot 2). The project developed an open-source toolbox within the statistical environment “R” enabling transparent procedures for raising sampling data and calculation of estimates of precision. The project further developed a substantial module for exploratory analysis of sampled data. Another main achievement of the project was the further development of the RDB-FishFrame data format into the standard data exchange format which is presently used by the RDB. Dissemination of experiences gained in the project was supported by a hand-on workshop in 2010.

The outcome of COST was a software product, entirely developed during the course of the project, which will necessarily need further adjustments and debugging. The effective usage of any tool must now be implemented and accompanied, in order to avoid institutes being left without help and, on the other hand, developing their own version of the software. The RCM did thereby in 2009 and 2010 recommend the onset of a COST 2 project addressing those issues as well as adaption of some codes to statistical sound sampling designs. The proposal for a COST 2 project from the RCMs was never picked up by the Commission. Since the termination of the COST project the codes have mainly been maintained on a national level by IFREMER, France. RCM NS& EA (2011) considered that a continuation of the COST project had tremendous potential in particular linked to the development of a regional database. Integration of COST tools within the RDB have been included in study proposals for the development of the regional database that repeatedly have been put forward by RCM NS&EA (2012, 2013, 2014, 2015) and the Liaison Meeting.

4.3 WebGR

WebGR (Web services for support of Growth and Reproduction studies) is a set of open source web services and tools developed within an EU project in 2008 (FISH/2007/07 Lot 1). The services support the organization and data analysis of age and maturity calibration workshops and are implemented on a website. They are therefore essential tools for quality assurance and coordination of biological sampling and analysis. The website consists of a repository of images, a set of web forms to run a calibration exercise online, a reporting module with the most common statistical analysis and import/export modules to manage images and results. The results are extracted in a standard format. Several workshops and exchanges have used WebGR since 2010. The tool has however not been further developed since 2010 which has resulted in problems and underutilization of its potential. The service is presently freely provided at http://webgr.azti.es but without any warranties in case of problems (since there are no funding available). In 2013 and 2014 the RCM NS&EA recommended for a collaborative study of improvement of WebGR. ICES PGCCDBS strongly supported such a study proposal.

5 Regional coordination activities - major achievements

The last 12 years of regional coordination meetings have been a slow but steady process improving coordination between MS. Reading through the RCM NS&EA reports it is evident that two main elements have contributed considerably to the development and way the RCM is working; the uptake of “the fleet activity matrix” (later appendix IV in 2010/93/EU) in the legislation and the implementation of the regional database.

The “fleet activity matrix” is a comprehensive way to overview all the fishing activities in the region. It gave MS a common way to structure and identify fisheries/métiers and a generic way to name them. This allowed for regional overviews of fisheries and their importance in terms of catches, effort and value. A generic nomenclature also allowed for regional overviews of sampling coverage and relative contribution of sampling by different MS.

A common understanding and nomenclature gave the RCMs the possibility to work towards sampling plans on the regional scale but no effective means to do so. Several days of meeting time were spent just on simple overviews on regional fisheries as data was accessible through simple excel sheets. Even if MS had agreed on a generic nomenclature there were always slight differences in how métiers/fisheries, fishing grounds, harbours and even species were expressed. This led to the uptake and implementation of the regional database in 2011.

The implementation of the regional database made the work in the RCM NS&EA far more effective as the meeting could analyze regional data instead of just compiling them. It also opened up a necessary discussion (reflected in different recommendations from the RCM) on harmonization across MS in coding of métiers/fisheries, harbours, areas and species. The RCMs have worked intensively, together with MS, the RDB steering committee and ICES to establish reference lists for all these elements.

The RCM NS&EA have also served as a platform for experts and NCs to exchange views, discuss and sometimes agree on, an array of present and foreseen topics of data collection. This has most likely built trust and common views on how data collection schemes should evolve in the future. The consultation process within fishPi, for example, shows a substantial level of agreement on the movement towards regional sampling plans.

Work carried out within ICES has also contributed considerably to regional cooperation in data collection. A key point is the increased awareness of the need for statistically-sound sampling instead of *ad-hoc* sampling. ICES groups such as WKPICS, SGPIDS, WGRFS and WGCATCH have developed an important and well-documented body of knowledge on fishery sampling design, implementation and analysis. There has to a considerable extent been overlap in expertise between these ICES methodological groups and the RCM NS&EA ensuring good communication. The ICES WKRDB, the RDB-SC and fishPi have developed the RDB data exchange format to meet requirements of statistically sound sampling.

A considerable amount of work has further been completed within the ICES framework to establish international protocols and standards for fish ageing and maturity determination. This work is nowadays carried out by ICES WGBIOP.

Major achievements by regional coordination activities are:

- A common understanding of the fishery activities on the regional scale
- Common nomenclature for métiers/fisheries, ports, species and areas and agreed reference lists
- Implementation of the regional database
- Trust building between MS
- Increased awareness on the need for statistically sound sampling schemes. Body of knowledge developed.
- Establishment of networks to improve quality assurance in age and maturity and maturity readings within and between MS.
- Improved efficiency in sampling activities through setting up bilateral agreements in the region.

6 Regional coordination activities – areas for improvement

Despite steady progress in the field of cooperation between MS within the RCM and ICES framework there is still little cooperation in the field of actual commercial fisheries sampling. The bilateral agreements in place are a good start, but the step towards truly regional sampling plans is still a big leap. There are several potential reasons for this.

A main point is the awareness of the need for statistically sound sampling schemes to get transparent, reliable and quality assured estimates to end-users. From a regional cooperation perspective this does however imply that regional cooperation for actual sampling moves from the concept of simple task sharing in sampling to a need for integrated regional sampling designs. The latter requires a much stronger theoretical background and integrated approaches to protocols, quality assurance procedures and data management. This has of course had an impact on the speed of the cooperation process within the RCM. The entire fishPi project was designed to enable this transition. It is however important to realize that development, implementation and management of such schemes will require constant monitoring, and input from statisticians will be needed long term. This might to some extent be solved when RCMs move to RCGs, as suggested in the recast of the data regulation 199/2008. It does however require that MS have experience and resources to prioritize regional work. MS and institutes show, in the fishPi questionnaire, willingness to do so but are in some cases limited by resources.

A second point is that there has been no clear road map on how to achieve regional sampling plans. This is important as there are so many different groups (RCM, ICES, STECF EWG) dealing with different aspects of regional coordination and regional sampling plans that it is difficult for individual MS or experts to get an overview of what is going on. Different processes are further dependent on the success of other processes. The RCM NS&EA 2013 did attempt to create such a road-map but concluded the following year that the road map was hampered by the absence of the new legislation, the lack of development of the RDB and the lack of establishment of the RCG process. Establishment of a robust and agreed road-map would preferably be a task for the RCM.

A third point is the absence of long term funding to meet regional needs. Tools, such as COST and webGR, developed to support quality assurance on the regional level have typically been funded through EC study contracts. This has been beneficial as the tools have been developed, but it also implies that there are no resources for maintenance or further development when the contract terminates. In the long term this means that the tools, essential for regional coordination, are not adapted to meet new requirements from e.g. science or IT systems. The consequence is that they are not utilized to their full potential or that they turn into national tools as development continues within individual MS. RCM NS&EA has repeatedly recommended continued funding for some essential projects but without result.

A key aspect is the lack of funding for the regional database (RBD) identified by the RCM NS&EA as a crucially important tool for development and maintenance of regional sampling plans. RCM NS&EA has repeatedly stressed the need for funding of development to meet their needs but up to now with no results. The RCM NS&EA 2015 expressed in the strongest terms “that the short term needs of regional cooperation were dominated by the overwhelming need to fund work related to the RDB, emphasizing that the RDB was not simply a data base, but also a means of facilitating the data analysis, skills, dissemination of best practices, and harmonization of work involved in regional data collection and estimation”.

References

ICES 2002 - 2014. Annual Reports of the Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS). See ICES website http://www.ices.dk/publications/library/Pages/default.aspx#k=PGCCDBS

ICES 2007a. Report of the Workshop on Sexual Maturity Sampling (WKMAT). ICES CM 2007/ACFM:03

ICES. 2007b. Report of the Workshop on Discard Raising Procedures (WKDRP). ICES CM 2007/ACFM:06, 55pp.

ICES 2008a. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU). ICES CM 2008/ACOM:32, 41 pp.

ICES 2008b. Report of the Workshop on Maturity Ogive Estimation for Stock Assessment (WKMOG). ICES CM 2008/ACOM:33

ICES 2009a. Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE), ICES CM 2009/ACOM:40, 43 pp.

ICES 2009b. Report of the workshop on sampling methods for recreational fisheries surveys (WKSMRF).

ICES CM 2009/ACOM:41

ICES 2010. Report of the Workshop on methods for merging métiers for fishery based sampling (WKMERGE), ICES CM 2010/ACOM:40, 94 pp.

ICES 2011a. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, ICES CM 2011/ACOM:52

ICES. 2011b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS).
ICES 2011 /ACOM:50

ICES 2012 onwards. Annual reports of the Working Group on Recreational Fisheries (WGRFS). See ICES web site: http://www.ices.dk/publications/library/Pages/default.aspx#k=WGRFS

ICES 2012a. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS).
ICES 2012 /ACOM:50

ICES 2012b. Report of the second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes. ICES CM 2012/ACOM:54

ICES. 2013a. Report of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS3). ICES CM 2013/ACOM:54.

ICES 2013b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2013/ACOM:56

ICES 2014. Report of the Working Group on Commercial Catches (WGCATCH). ICES 2014:/ACOM:36.

ICES 2015. Report of the Working Group on Commercial Catches (WGCATCH). ICES 2015:/SSGIEOM:34.

Jansen, T., Degel, H., Vigneau, J and. Jardim, E. 2009. Definition of Standard Data-Exchange Format
for Sampling, Landings, and Effort Data. ICES CRP No. 296.

Nedreaas, K., Stransky, C., Jardim, E. and Vigneau, J. 2009. Quality assurance framework—the concept of quality assurance applied to fisheries data and its operationalization under the ICES scope. ICES CM 2009/N:06.

RCM NS&EA 2014 -2013. Report of the Regional Co-ordination Meeting for the North Sea and Eastern Arctic (RCM NS&EA) See website https://datacollection.jrc.ec.europa.eu/docs/rcm

ANNEX 2 WP1 Work programme towards regional sampling plans

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION IN FISHERIES DATA COLLECTION"

Deliverable 1.2 – Work programme towards regional sampling plans

Lead: SLU, Sweden.

Core Team: Katja Ringdahl, Maria Hansson (SLU, Sweden)
Mike Armstrong (CEFAS, UK England)
Jose Rodriguez (IEO, Spain)
Jörgen Dalskov (DTU-AQUA, Denmark)
Sieto Verver (IMARES, Netherlands)
Christoph Stransky (TI-SF, Germany)
Joël Vigneau (IFREMER, France)
Els Torreele (ILVO, Belgium)
Liz Clarke (MSS, UK Scotland)
Estanis Mugerza (AZTI, Spain)

Table of Contents

1 Background and context

1.1 What is a regional sampling plan?

The EU Data Collection Framework (DCF) and Control Regulation require Member States (MS) to collect a wide range of specified data from fisheries and stocks in defined regions, to support implementation of the Common Fisheries Policy (CFP). These data are used to meet the needs of end users such as ICES, Regional Fisheries Management Organizations (RFMO) stock assessment working groups or other bodies providing advice on fisheries management. In most cases these end users need data aggregations and estimates that relate to fisheries at a regional scale.

Some data are exhaustive (provided by all fishers), such as landings and effort for commercial vessels required to keep EU logbooks. Other data are provided by fishers in part, such as some information on small scale fisheries with less stringent logbook requirements. Other data, not provided by fishers, require the use of statistical survey methods to generate estimates of the parameters of interest. The latter includes the species, age and length composition of different fractions of catches, quantities discarded, data on small scale and recreational fishery catches, and data on the by-catch of non-target species such as mammals and birds. Additionally, data on individual fish for the estimation of biological parameters such as growth, maturity, sex ratios, stomach contents etc. are also collected using statistical sampling methods either from commercial vessels, dedicated research vessel surveys or other sources. Statistical sampling plans are needed for the collection of all non-census data to ensure that sampling methods are appropriate, statistically robust and deliver the data and estimates required.

At present, EU Member States devise their own sampling plans to deliver what is expected from EU legislation. The national sampling plans have often been adapted from schemes and targets in place before the DCF and were driven more by meeting DCF sampling targets rather than by sound statistical design. As a result, origins data collected at a regional scale are from a diverse collection of national sampling schemes that may or may not conform to the type of statistical standards given by EuroStat or equivalent guidelines, and are likely to provide poorer estimates and value for money than could be achieved by a better coordinated and more statistically sound approach to regional sampling. The transparency of national sampling schemes and demonstration of data quality also currently remains poor in some cases.

A regional sampling plan would, in contrast, be designed from agreed objectives based on end-user needs with the outset to provide data on the collective activities of nationally operating fleets that are fishing on shared stocks in regional seas. By regional seas we mean the large ecoregions such as the Baltic, North Sea and Western Atlantic.

Important elements of a regional plan would be:

- Agreed objectives based on end-user needs
- Sampling design and estimation methods
- Selection protocols
- Sampling protocols
- Integrated quality assurances framework and common tools for quality checks
- Regional database for regional storage and management of data, including common software tools for analysis and estimation.

The operation of regional data bases as the means of housing the collected national data sets, and facilitating the analysis and estimation at national and regional level is a key element in this regional approach.

The fishPi project shows how both commercial and recreational fishery sampling could be better coordinated at a regional level, and illustrates the potential benefits in terms of effectiveness, data quality and statistical robustness. It is a first step in a process using simulations of regional sampling schemes based on actual data collected in recent years.

1.2 Who are the end users and what are their needs?

End users are defined in the DCF Regulation as bodies with a research or management interest in the scientific analysis of data in the fisheries sector (article 2 of the DCF Regulation 199/2008). For the purposes of defining regional sampling plans, the end users can be considered to be advisory bodies such as ICES, STECF (Scientific, Technical and Economic Committee for Fisheries) or RFMOs. These end users require a variety of data and estimates on a recurrent basis at a range of aggregation, from detailed (anonymised) data to highly aggregated data, to develop advice on fisheries management. End users at a national level, can be a variety of national governments, local bodies, NGO’s and others, all with a clear interest and legitimate requirement for fisheries data.

The regional fishery sampling schemes will typically need to deliver at least the following types of estimates of fishery-based and stock-based biological variables:

- Catch compositions such as the age and length compositions by species, and the species composition of mixed landing categories and discard components of the catch. For mixed fishery models, catches and catch compositions are needed at the scale of the fisheries defined in the analysis.
- Catches, where these are not recorded exhaustively and must be estimated from sampling surveys (e.g. discards; catches of some small-scale commercial fisheries and recreational fisheries).
- Biological data to estimate sex ratios, maturity ogives, growth parameters where these cannot be obtained from research vessel surveys.
- Data to estimate the impact that fishing activities have on the marine ecosystems (bycatches of marine mammals, sea birds and turtles, stomachs)

Typically all these estimates are required for various subsets of the population, most noticeably stocks or species, temporal periods – usually annually or quarterly, métiers, areas, and so forth.

ICES currently estimates parameters such as total regional landings and discards at age or length by species and fishery, using procedures within its InterCatch database to combine national estimates that are submitted following data calls. These final regional estimates for commercial fisheries are typically provided only with basic quality indicators such as numbers of samples. There is no equivalent system for recreational fishery survey data or data on by-catches of protected endangered and threatened species (PETS).

1.3 Sound statistical design and regional sampling plans

The regional design, given the complexity of fisheries sampling, will typically be a stratified multistage design utilizing sampling frames for the selection of sampling elements and employing probability based selection mechanisms. Within a regional design individual nations could be considered as strata, thus maintaining an element of national autonomy in the collection of data. It is envisaged that regional designs can, if necessary, operate alongside existing national sampling designs.

The regional sampling plan, and its possible optimization, appropriate for commercial fisheries will follow the general scheme shown in Figure 1. The design is driven by the estimates needed by end users, and follows well established methods of probability based sampling.

Designing appropriate sampling schemes will be a step-wise iterative process requiring close cooperation and consultation at each stage, both between and within the institutes carrying out the sampling, and also with end users.

Statistically sound sampling schemes require probability-based selection of sampling units from the available units, and that the coverage of the population, non-response rates and refusal rates are calculated. Appropriate estimation methods take account of the design and sampling hierarchy inherent in the sampling scheme, weighting the data according to their probability of being sampled. The sampling design developed is appropriate for the defined parameter estimates from the defined study population. Additional requests for data potentially require the redevelopment of an appropriate sampling scheme.

The key steps in the development of a regional design are as follows:

- The definition of the parameters to be estimated and the target population for which these estimates are required to meet end-user needs– typically these are the total annual catch for which discards, age or length compositions, sex ratio etc need to be estimated. The study population would most likely be defined in terms of fleets of certain types of vessels which fish certain species. Given the overlap between fisheries and stocks, the process will of necessity be a step-wise process developing expertise by focusing on certain segments of the overall population (e.g. pelagic fleet, fleets catching the majority of the demersal species, fleets fishing within a certain region).

- *Construction of appropriate sampling frames.* These are the lists of sampling units that can provide access to all of the elements of the target population, such as catches of individual fishing trips. In a regional sampling plan, the frames will encompass the catches of all countries in the region and need to be carefully defined to cover as much of the target populations as possible. Separate frames can be set up for components of the design that are sampled independently, such as at-sea and on-shore components.

- *Choice of sampling stratification.* Sampling strata are non-overlapping subsets of the sampling units in the sampling frame, which ideally group together homogeneous sampling elements. As effort allocation can be varied between strata this enables the precision of estimates to be improved. Criteria for defining sampling strata can be any recognizable trait of the sampling units which reflect homogeneity in the data collected. Typically this might be vessels grouped by size, typical fishing gear or trip duration in the case of at-sea schemes. For on-shore schemes it could be markets where particular fish species are sold, the port size, or specific handling or landing practices.

- *Effort allocation.* This is the specification of the number of sampling events for the whole scheme and their distribution between the sampling strata. The effort allocation will be influenced by the variability in the population being sampled and a desired precision of estimates, and inevitably will need to take into account the available resources. Allocating most effort to the most variable strata is common practice where there is a desire to obtain precise estimates. Determination of stratification and effort allocation is often carried out through simulation studies.

- *Identification of appropriate probability based selection methods.* In the first instance this is the random selection of the sites and day, or a vessel, from the appropriate sampling frame. The sampling designs employed for fisheries sampling invariably then rely on multistage sampling; e.g. the selection of a vessel in a port, the selection of size classes from that vessel's landings, the selection of fish from a size class, in order to obtain the individual fish from which measurements are taken and otoliths collected. At each stage the elements have to be selected appropriately and the sampling probabilities recorded. The importance of recording the sampling probabilities at each stage in the sampling hierarchy is critical, as this determines the weighting the sample receives in its contribution to the final estimate. The use of simple random or other probability based selection methods ensures that in the ideal situation the estimates will be unbiased and that statistically valid measures of variation can be calculated. In the less than ideal situation often encountered in fisheries sampling, recording the selection process enables sources of bias such as non-response and refusals to be detected to potentially be corrected. Employment of estimators that respect the sampling design and the sampling hierarchy are used in the multistage sampling.

- *Optimisation of sampling schemes.* The simulation models can be employed to explore alternative sampling designs that differ in stratification and effort allocation. These simulation models can identify how sampling effort could be re-allocated within a region to provide better sampling coverage and improve precision. The development of methods to optimize sampling across species and fisheries, for more complex estimates such as age and length compositions, will require considerable research and development that will best be achieved in stages using subsets of data such as in the fishPi case studies, and in close cooperation with Member States and end users.

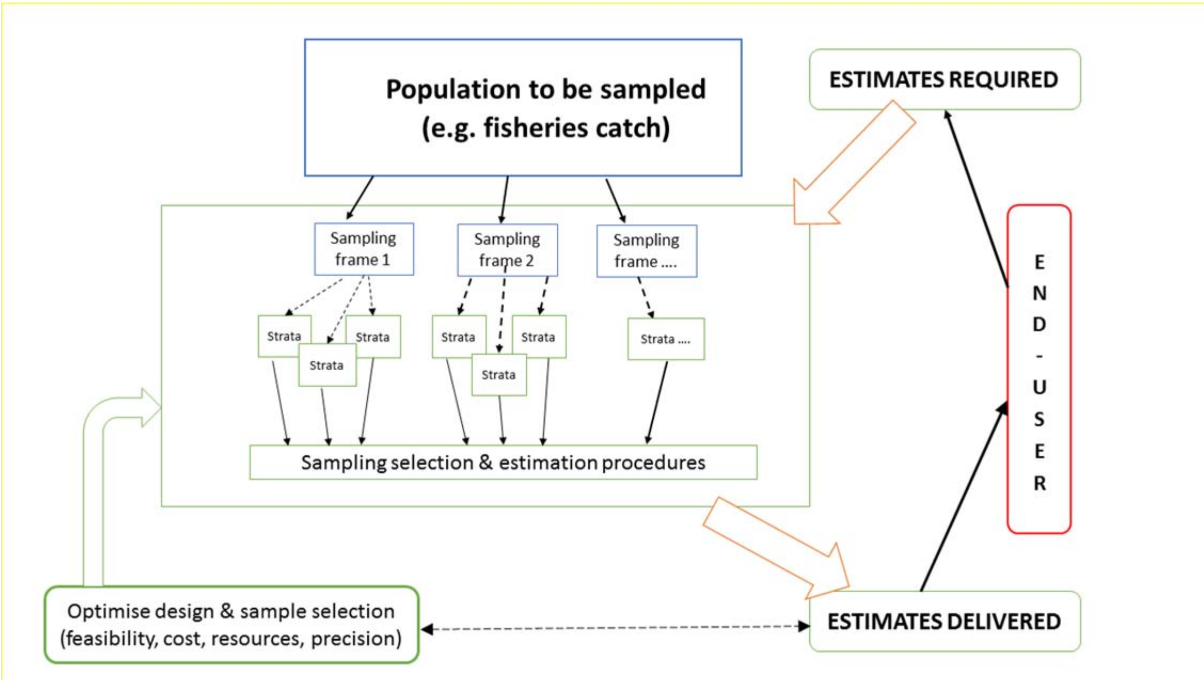


Figure 1. Overall schema for a regional fishery sampling plan and its optimization. This general scheme can be applied to commercial and recreational fisheries although the details will differ.

2 A regional work programme towards regional sampling plans

A regional sampling plan includes: agreed objectives based on end-user needs, an integrated regional sampling design, standardized sampling protocols, a common approach to quality assurance, and regional tools for the management and dissemination of data. All these different aspects need to be developed before a regional sampling plan can be implemented. The fishPi project (WP2-4) has contributed significantly to several of the elements. It is also more than likely that a regional plan will be constantly tuned after its implementation as fisheries and landing pattern change and science evolves. Management of the plan and its implementation are expected to be a task for the future Regional Coordination Groups (RCG). Different aspects and elements of foreseen sampling plans, including the fishPi contribution are described below. Short term (2016-2017) development needs, identified by other WPs in fishPi or elsewhere, are identified. It was not possible to make a timeline, since after the termination of fishPi there are no regional resources available for further development of such plans. Heads of Institutes and NCs have in the fishPi questionnaire shown willingness to contribute with expertise and staff time but do in several cases lack resources to do so. Prioritization needs to be completed by the relevant RCMs/RCGs.

2.1 Management and development of the regional sampling plans

Regional sampling plans are anticipated to be managed by the Regional Coordination Groups (RCGs). The mandate and tasks of these groups are described in the recast of DCF (199/2008). The groups are an evolution of the Regional Coordination Meetings (RCMs) that have taken place since 2004. The role and set- up of these groups have been elaborated in various expert groups (STECF, RCM), where it has been identified that regional sampling plans require substantial cooperation between MS during and between meetings. Subgroups working on identified topics to, for example, improve sampling designs and methodologies on the regional or supra regional level, may also be the best way to utilize human and financial resources on the regional scale. A possible set-up for an RCG is shown in figure 2.

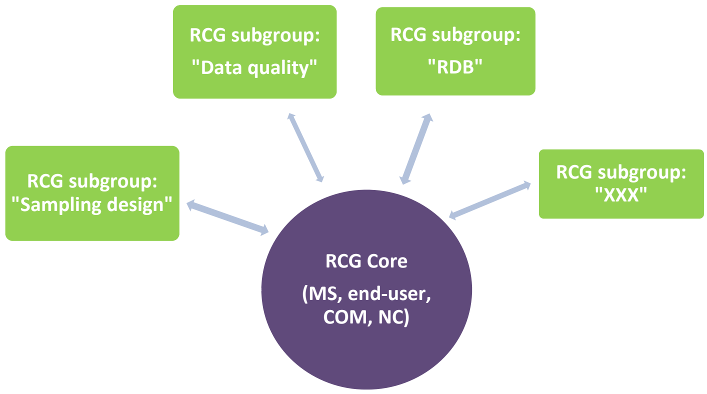


Figure 2. A potential set-up for an RCG (RCM Baltic 2015).

There are several long term objectives identified for the regional sampling plans, and the consultation procedure withinWP1 of fishPi highlighted a strong and unified support towards a regional approach and regional sampling plans

Long term objectives for regional sampling plans include:

- To more clearly align with the appropriate end-user needs, and be fully transparent in terms of design, data held, data quality and methods of analysis to provide regional totals.
- Adoption of probability-based (statistically sound) sampling schemes and estimation methods aimed at providing representative data. This is a prerequisite to minimise bias, provide reliable estimates of precision, and allow optimisation. The approach taken will be in line with EuroStat and other appropriate guidelines.
- Critically evaluate the feasibility of these regional plans given the requirements to provide the highest quality of data for as many stocks and fisheries at a regional scale as is possible within resources available, and taking account of national requirements, available skills, and any other relevant factors affecting flexibility to adjust sampling.
- Ensure a high level of quality assurance of archived data held in national and regional data bases.
- Make best use of sampling and analytical skills and resources available across the countries with fisheries in a region, including training and task sharing.

The development of regional sampling designs will be an iterative, step-wise process requiring close co-operation and consultation at each stage between sampling design experts and expert groups, institutes carrying out the sampling, and a range of end-users. The overall process should be coordinated by the new Regional Coordination Groups, which would require the mandate and sufficient skills and resources for this. Sufficient resources also need to be allocated in national work plans for development of the regional sampling plan. In the consultation process provided by WP1 institutes and MS were willing to support the regional work with staff time. As resources in several cases are limited some respondents replied that this would need to be completed on a case by case basis.
An important consideration is the steps that would be required in the coming years to move towards the goal. fishPi supports this process by providing a statistical framework for the sampling schemes, including the sampling design principles, associated estimation methods that reflect the sampling design and hierarchy, and protocols for the selection of sampling units. The case studies developed in fishPi WP2 provide pilot studies for the development of regional sampling plans, testing out the approaches described above, including the development of a format and agreement for data-sharing

for this purpose. The case studies for commercial fisheries found that a regional approach to effort allocation was likely to be more efficient than the current national allocations.

2.1.1 Short term plan

The short term development plan needs to involve:

- RCMs to evaluate the expertise and resources needed for the Regional Coordination Groups to develop regional sampling plans.
- RCMs to identify a process of collaboration with end users, Advisory Councils, and expert groups to help in development of elements of the sampling plans.
- Further work in a new project or under the umbrella of the RCGs to further refine regional sampling designs.

2.2 Regional designs of sampling plan for commercial fisheries

Several types of regional sampling plans are potentially needed to collect data appropriate for stock assessments and fishery management advice and other end user needs.

Fisheries are complex and there is overlap between fishing fleets in terms of species caught, gears used and regions fished. Furthermore, accessibility of the catch changes between fleets and different countries have different methods and limitations for accessing and sampling the catch. Detailed knowledge of the different fisheries and sampling methods for the different countries needs to be built up at a regional scale through close collaboration between members of the relevant institutes, to enable appropriate regional sampling designs to be developed.

In many ways regional sampling plans are most clearly suitable for shared stocks and mixed species fisheries where data on a number of important species and stocks can be collected using a single sampling design. Here species groupings with taxonomic and ecological similarities are important to consider, as are the spatial and temporal distribution and activities of the fishing fleets that target these groupings.

Within fishPi WP2, considerable effort has been utilized to simulate different types of stratification within regional sampling plans within 4 case studies. Their effectiveness has been tested as well as implications for Member States in terms of changes in sampling commitments. The case studies were:

- Pelagic species in the North Sea and western Atlantic
- Southern North Sea Flatfish fisheries
- Northern North Sea demersal fisheries
- Northern and southern Hake stocks

The case studies have however revealed that more work needs to be done before such designs can be implemented in real life. This is supported by the consultation procedure (WP1) in which responding institutes and MS highlighted some logistical constraints (fishP WP1 Annex 3) . A key element that needs to be resolved is the decision making process in the Regional Coordination Groups.

2.2.1 Short term plan

The short term development plan needs to involve:

- A scientific planning group for key fisheries should be identified and established to oversee the development of the regional sampling schemes. This will be a step-wise iterative approach, which will need to focus on sections of the fleet such as pelagic or demersal trawlers, and will need to consider the provision of data for the stocks fished by these fleets.

- The results of the simulation studies for commercial fisheries have indicated implications of regional sampling designs. More research is however required to improve the evidence for the move to regional sampling schemes, and to determine the appropriate effort allocation to meet agreed objectives. This will be a step-wise iterative process involving collaboration between all relevant parties. In the short-term, research into sampling designs would include:
    - Improved information of current sampling schemes and other improvements to the population data – this would require a new data call.
    - Extending the single species case studies to multi-species sampling designs – this could require a new data call.
    - Similar simulations using models of discard weights by fleet.
    - Similar simulations using biological data or models of discard weights by fleet. These simulations are likely to be on a single species basis so the analyses would need to be repeated for several species.
    - The inclusion of random species selection.

- Discussions between participating Member States (and institutes) to resolve the issues highlighted in the simulation studies before a regional sampling design could be introduced. Negotiations in regional bodies such as RCGs or survey planning groups is required to resolve these issues, which include:
    - The regional allocation of effort would require redistribution of effort from one country (or fleet) to another.
    - The sampling of landings from foreign vessels.
    - The collection of biological data for species of less importance to the country.
    - The mechanism for inclusion of national requirements in a regional sampling design.
    - The appointment of groups that focus on the design & implementation of a certain fisheries sector in a long-term setting.

2.3 Regional sampling plans for estimating impact of fishing activities on the marine ecosystems, small scale fisheries and recreational fisheries

Current data on the bycatch of protected endangered and threatened species (PETS) and stomach contents for fish species are not covered by the DCF. This means that sampling presently is not mandatory for MS or institutes under their national plans. Coordination of data collection of these variables has consequently not been discussed and no analysis undertaken, at the different RCMs. The likely inclusion of these variables under the new EU MAP (Commission Implementing Decision of [Pending] adopting a multiannual Union programme for the collection, management and use of data in the fisheries and aquaculture sectors for the period 2017-2019), currently under discussion by the Committee for Fisheries and Agriculture, could have big implications, both from a technical point of view and a cost perspective.

Sampling plans for PETS

The monitoring schemes in operation under the current Data Collection Framework (DCF) can contribute to the assessment of by-catch of cetaceans and other species. They are however not sufficient on their own as currently implemented by MS since the emphasis of at-sea observer programmes is focused on discards. Not all fisheries are adequately covered and many issues, including design and sampling protocols would need to be modified and/or extended if the new EU MAP monitoring is to be the sole source of information. Monitoring under Regulation 812/2004 is much more specific for cetaceans, and has included the use of dedicated observers and remote electronic video recording.

Two approaches to the sampling of PETS have been explored through case studies within fishPi WP3. The approaches can be further developed within, for example, the RCGs.

The first, risk based assessment (fishPi WP3 Annex 15), would initially identify areas and fisheries where by-catches of PETS species are considered to be causing a conservation threat, and document the sampling coverage of these fisheries under the current DCF by different MS observer programmes. A selection of métiers and areas can be used as an example for the practical implementation of coordinated sampling for bycatches of PETS. In the selected métiers and areas, the following tasks would be completed:

- Characterize the fisheries taking into account the effort, catches, seasonality, etc.
- Identify MS involved in these fisheries.
- Document current sampling coverage with the at-sea observer programmes under current DCF.
- Analyze possible coordination among MS to improve the sampling.
- Analyze problems and improvements in the practical implementation.
- Propose alternative sampling methods (CCTV, questionnaires etc.)
- Identify potential cost implications in the monitoring of these variables.

A second approach would be the sampling of by-catches of PETS fish species. This case study (fishPi WP3 Annex 15) has used real data from logbooks and national observer programs and aimed to identify a general set of steps to set up a regional program for sampling bycatch of protected, endangered and rare species (PETS). It differs from the approach above in the type of PETS targeted; this study focuses on small to medium sized fish and elasmobranch species that may be encountered in the catch fraction commonly handled by at-sea observers. Species-specific fishery descriptions based on historical logbook data can be used to identify potential stratification with regard to seasons and areas. Based on the steps identified, different sampling scenarios can be examined. The compiled data from commercial fisheries can be raised to regional level to approximate the potential sampling effort needed given actual fishing patterns and effort in the selected region.

Sampling plans for stomach contents

Historically, different research surveys have been used to collect stomach content data but it has never been the main objective of these research surveys. Again, specific sampling programmes to collect these data needs to be modified and/or extended to provide surveys to cover all the needs of relevant end users, new protocols, etc. Furthermore, in the case of stomach contents there is also a need to consider the analysis of them.

Defining a sampling plan using a single stock as a case study has no added-value as it does not provide enough information to show accurately how MS could coordinate and adapt a multi-species regional sampling programme to support the move towards the Ecosystem Approach as required by

European policies. The main issues to resolve are e.g. how MS can build up sufficient, statistically-sound samples of all the key species needed for food web characterization and for quantifying predator-prey relationships and predation rates within different types of multispecies and ecosystem models.

With the aim of encouraging collaboration at a regional level and to propose common sampling protocols, a case study using the Bay of Biscay and Iberian waters has been selected as an appropriate eco-region to tackle stomach data collection integrated across three MS. The sampling schemes for collection of stomachs, and data availability, are very different between these MS, providing a good example of how different collection protocols could be applied and what form of coordination and cooperation would be needed to ensure that the data are compatible and can be combined into a statistically robust regional ecosystem data set for analysis.

Sampling plans for recreational fisheries
Regional coordination of recreational fishery surveys is needed to ensure that end users are supplied with the catch estimates or other data needed by them, at the required spatial resolution and temporal coverage, and with the required quality. Coordination is a role for the lead scientists of the surveys in each country, the Regional Coordination Group, and the ICES Working Group on Recreational Fisheries Surveys (as technical expert advisory group).

The appropriate design of recreational fisheries in a region does not necessarily have to be completely harmonized between countries. The most important attribute is that surveys have robust statistical designs to minimise bias and allow correct calculation of precision. Catch estimates from different surveys can then be combined. Information must be collected to allow evaluation of the potential for bias, for example non-response rates, refusals and incomplete coverage.

Given the availability of human and financial resources for fishery sampling in each country, there will be a need for a trade-off in sampling effort allocation between large scale and small scale commercial fleets and also recreational fisheries where data are required. The same is true for different types of PETS. There will never be a single optimized regional sampling programme, only a set of scenarios which need to be evaluated. To do this objectively, the types of analysis developed within fishPi and applied in the case studies in WP2 are essential to show the impact of altering the national allocation of sampling on the quality of estimates for each species and stock for which data are needed for assessment and management

2.3.1 Short term plan
The short term development plan needs to involve:

- As a first step substantial work needs to be carried out between the main end users, data collectors and experts in sampling, under the umbrella of the RCGs to identify:
    - Priorities on data needs and criteria (i.e. set by-catch limits needed for designing monitoring)
    - Standardize guidelines and protocols
    - Monitoring methods taking into account cost/benefits analysis
    - Practical implementation considerations
    - High risk by-catch métiers by region in case of PETS
    - For small scale fisheries and recreational fisheries evaluate national requirements for data to support national inshore management schemes, and find a trade-off between regional and national needs.

- Second step: RCGs designing regional plans to cover needs identified in the first step.

The collection of these new variables implies:

- Important intersessional work between end-users and data providers. RCGs should push and lead this work. Probably some funding and resources will be needed.
- Expertise and training as new methodologies will probably be used to collect and analyze these new variables (i.e. CCTV in case of PETS, genetics for stomach content).
- Expertise and training to implement the principles of probability based sampling to these new variables (PETS, stomach) and fisheries (SSF, RFS)
- Increase in human resources
- Essential to have a regional data base, to upload collected data in a harmonized, standardized and transparent way. Responsibility of RCGs and RDB-SC.

2.4 Agreed protocols and methodologies

The sampling of commercial marine fisheries at a regional level will require common manuals to define the key principles and protocols for effective and comparable sampling of catches and standardized recording of the data collected. The manual should contain specific details and references to enable anybody involved in data collection to find all the information required to conduct all the different types of sampling described. The manual should be reviewed on a regular basis as appropriate and updated whenever the need for a significant change is determined. A new version number of the manual should be then published in a common SharePoint.

National institutes in charge of the sampling will need to ensure that survey staff familiarize themselves with the manual and follow the protocols during each sampling event. To address the non-standard situations and particularly challenging sampling situations often encountered in fisheries data collection, protocols can be developed and disseminated between Member States, thereby facilitating the adoption of best practices across the region. The protocols covering the steps from data collection to estimation would need to be aligned within the region to allow consistent approaches to quality assurance and estimation.

Many handbooks and guidelines are already well established in the data collection community, for example, handbooks for maturity staging and the determination of fecundity and egg development. Also sampling protocols for research vessel surveys are well embedded in general and developed within the dedicated expert groups. For commercial on-shore sampling as well as commercial at-sea sampling, no widely accepted sampling protocols are yet established. On a national level, (as shown in e.g. PGCCDBS 2012) guidelines exist on various topics.

Various ICES expert groups (PGCCDBS, WKPICS, SGPIDS, WGRFS, WGCATCH) have, over the years, highlighted the need for standardized sampling protocols, common reference lists and specifications of best practice for fishery sampling. These are also a prerequisite for successful regional sampling designs for the collection of data for management purposes

fishPi are supporting this work by providing the statistical framework for the sampling schemes, including the sampling design principles, estimation methods which respect the designs, and protocols for the selection of sampling units. Respondents of the fishPi consultation procedure were generally positive to adoption of regional data collection protocols provided they do not jeopardize data collected in support of national and regional policy objectives.

2.4.1 Short term plan

The short term development plan needs to involve:

- List current, complete and agreed guidelines and ensure they are available and embedded in MS.
- The current list of protocols should be compiled and reviewed at the RCGs. The RCG should also facilitate the dissemination of the guidelines through a dedicated website, ensuring public access to the guidelines.
- The RCGs should (intersessionally) set up a list of required protocols that need to be completed and adopted prior to the implementation of a regional sampling plan. Ideally, this intersessional work spans multiple regions to ensure similarity in approach, priorities and synchronization of methodologies.

2.5 Regional data bases and the management of data collected through regional sampling plans

Regional databases (RDBs), in which regional fisheries dependent data are stored in common formats, are an integral component of regional sampling plans, ensuring transparency and consistent standards for data processing and dissemination across Member States. The RCMs consider the RDB a prerequisite to coordination of data collection on the regional scale. The RDBs also have the potential to support countries in analysis and estimation of national biological data. This means that the database has the potential to play a key role in the cost-efficient use of resources by member states to analyse and check data and generate estimates to answer data calls.

The RDB has been developed incrementally over several years and is now populated with data from ICES regions. In the RCM/RCG context, it is being used for regional co-ordination and quality control. It has been hosted by ICES since 2012 which has resulted in considerable improvements in terms of functionality and data consistency. There is a Data Policy document that has been agreed by the MS participating in regional co-ordination (except one MS) and EC lawyers agreed with the content of the document. This document also covers confidentiality issues. The data storage on the regional level is conducted by MS uploading their data in the standard RDB format, being defined in the exchange format document (http://www.ices.dk/marine-data/data-portals/Pages/RDB-FishFrame.aspx). The format is currently further developed, by fishPi and by ICES WKRDB, to meet requirements of the future regional designs and corresponding estimation procedures.

A generic overview on how the RDB will work has been compiled by the RDB steering committee and is presented in Figure 3 (more end-users than RCGs and ICES can of course be included). However the functionalities in the database need to be developed to meet requirements arising from a common approach to statistically sound sampling and regional sampling plans. This development needs to take a step by step approach following the development of sampling designs and robust estimation processes. To accommodate this need for flexibility, to ensure transparency, and to build on, rather than repeat work carried out by others, parts of the development can be developed outside the actual database (see figure 3). The most likely medium for these developments is the statistical software language "R".

The overall goal of data management, data storage, data handling and dissemination is to improve availability of quality-assured data and improve transparency and efficiency in the processing, management and dissemination of data. This is linked to an emphasis on regionalization of the data collection and dissemination processes as most end-users operate at the regional level. This will also ease the burden of data calls within the MS.

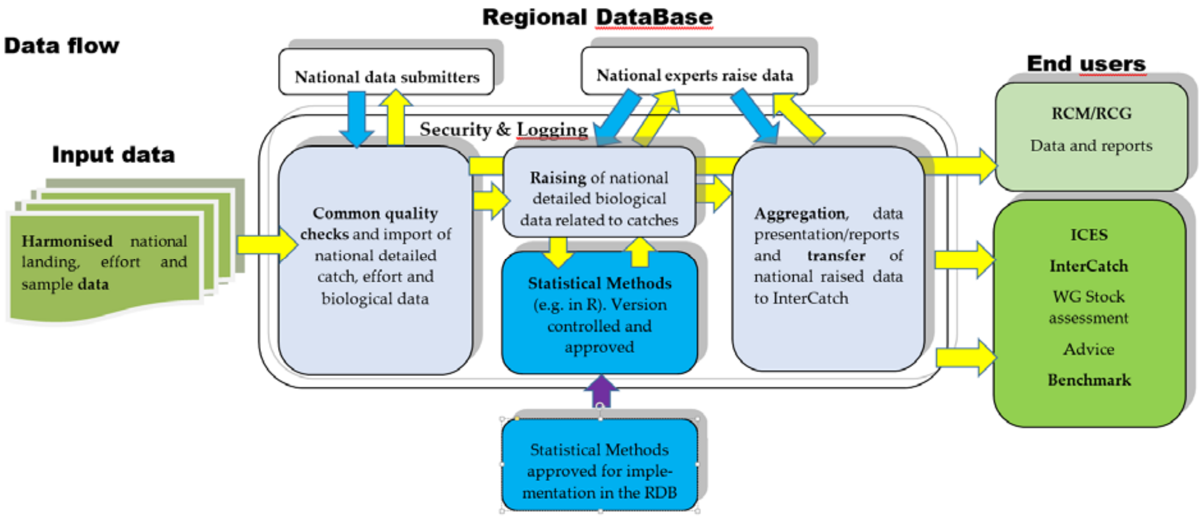


Figure 3. Future RDB system structure.

fishPi are supporting this work by providing the statistical framework for the sampling schemes, including the sampling design principles and estimation methods. These might in the future be adopted in the RDB framework. fishPi WP2 (fishPi WP2 Annex 9) has further continued to develop the format for sampling data proposed by WKRDB 2014. This format takes elements of sound statistical sampling into account and has the potential to be adopted for the exchange format for the RDB. fishPi has further developed functions, primarily for quality checking, using these formats which can be accessed from national and regional databases. fishPi developed a simple format for sharing of log-book/sales-note data for use in the simulation testing of sampling designs.

2.5.1 Short term plan

The short term development plan needs to involve:

- Further development of the Regional Database to meet requirements for future regional designs, quality assurance checks, estimation procedures and dissemination. The following additional development has been proposed by the Liaison Meeting and RDB Steering Committee:

  - Development of additional reports for analysis and data tabulation to support regional coordination
  - Development of additional reports for preparing Member States' Annual Reports
  - Extended data logging (What has been uploaded when)
  - Implement quality control functionality
  - Explore options and cost implications of implementing external tools (e.g. fishPi) by the RDB
  - Requirements and automation of data call procedures
  - Development of statistically sound raising procedures used by the RDB
  - Update of the existing data user roles and access module

2.6 Regional sampling plans and a Quality Assurance Framework

In the present system MS are responsible for the quality and validation of their own data, and must report on basic quality indicators (e.g. number of samples, number of measurements) as required by the DCF Regulation. The data validation is implicit and there is no reporting of which controls have been used on the national data. For assessment purposes, only limited data validation and quality checks can be done at a regional level, since the detailed data are not available. Each MS organizes their tasks nationally, trying to cope with all requirements from data collection to data transmission, with little collaboration between countries. However, numerous working groups (eg ICES WKPICS, SGPIDS, WGCATCH) have recommended on best practice in different fields of the DCF requirements, but national experts are often left alone in implementing these in their national context.

In the present system is it therefore difficult to assess quality of data and its impact on quality on the final assessments. End-user feedback is as a consequence mainly driven by data transmission failures, and less by quality assessment and improvement suggestions. Moreover, the end-users are often not aware of data being collected but not used (eg. maturity, discards, recreational fisheries).

The overall goal is to have a Quality Assurance Framework (QAF) for a regional sampling plan that covers the whole process from sampling design to data transmission with common and agreed quality checks, data validation and quality indicators. MS are responsible for the quality and validation at the national level, and a regional body (e.g. RCG) has the mandate to evaluate the quality at the regional level, and come back to MS if any issues need to be addressed. The end-user feedback is facilitated in a system where all detailed data is centralized at a regional level, and where adjustments and error corrections are being addressed comprehensively. Transparency between national datasets, regional datasets (RDB) and end-users' databases (e.g. ICES InterCatch) are ensured, both in terms of data availability and statistical procedures. A core team of dedicated experts on a regional level has the potential to bring the most advanced science to the whole system, helping each MS in capacity building and ensuring the best quality (e.g. consistency, harmonization, cost benefits) of the data being transmitted to end-users.

fishPi WP4 has proposed a list of quality checks, together with code details, to be used on national and regional datasets. Some of these quality checks are using the most advanced science in outlier detection. MS and Institutes conclude within the fishPi consultation process that they have sufficient national expertise to run and interpret codes such as the ones developed within fishPi. The quality checks have been built on the exchange format (fishPi WP2 Annex 9) and data structure proposed by the fishPi and ICES WKRDB. A calendar (fishPi WP4 Annex 18) for the implementation of quality checks at national and regional level is further proposed by fishPi, with the objective of having quality data timely for end-users.

2.6.1 Short term plan

The short term development plan needs to involve:

- The fishPi R package for quality checks are available in a public website (https://github.com/ldbk/fishPifct). A number of countries have volunteered to export their data into the fishPi data format and run the quality checks, in order to test the codes and report back to fishPi developers for improvement and potential new quality checks.
- ICES WKRDB 2016 should trial the data format and data structure.
- All MS should then be encouraged to export their data into the fishPi format (fishPi WP2 Annex 9) and test the codes developed in fishPi. They should then report to RCGs for feedback on their conclusions and suggestions for improvements.
- Quality checks to be implemented in the RDB and tested on regional datasets.

3 Conclusions

The fishPi project has significantly contributed to several elements needed for the development of regional sampling plans. These elements include a statistical background to regional sampling plans of commercial fisheries, different approaches to the collection of variables needed to estimate fisheries impact on marine ecosystems and quality indicators. However after the termination of fishPi there are no explicit regional resources available for further development of such plans. There are however, as indicated in the work programme several aspects that need to be further developed, fine-tuned, discussed and agreed.

Sufficient resources need to be allocated for development of the regional sampling plans. The consultation process within fishPi WP1 revealed that there is a substantial amount of expertise on different aspects of regional sampling plans available in different institutes and MS. There is willingness to support the regional work but resources in terms of staff-time and funding are often limited. It is important to realize that development of regional workplans is a step-wise iterative process involving consultation and collaboration with different parties ranging from data collectors, coordinators and main end-users. Robust resources to assure momentum are needed for several elements in the development of regional plans and need to be secured. Two key aspects are:

- EU to support a statistician team with time allocated to regional tasks.
- EU to support funding for further development of the Regional Database to meet requirements for future regional designs and estimation procedures.

The fishPi project
Regional co-operation in fisheries data collection

fishPi project (MARE/2014/19)

“STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION”

Deliverable 1.3 – The regional consulatioan process

Lead: SLU, Sweden.

Core Team: Katja Ringdahl, Maria Hansson (SLU, Sweden)
Mike Armstrong (CEFAS, UK England)
Jose Rodriguez (IEO, Spain)
Jörgen Dalskov (DTU-AQUA, Denmark)
Sieto Verver (IMARES, Netherlands)
Christoph Stransky (TI-SF, Germany)
Joël Vigneau (IFREMER, France)
Els Torreele (ILVO, Belgium)
Liz Clarke (MSS, UK Scotland)

Table of Contents

Executive summary

Member States (MS) have worked together since 2004 in Regional Coordination Meetings (RCMs) to coordinate national data collection programmes under the EU Data Collection Framework (DCF). In recent years it became clear that regional data collection plans would offer a number of benefits over non-coordinated national sampling plans. These benefits include more appropriate sampling designs, more rigorous statistical methodology, harmonised data formats and the adoption of standardised protocols for collecting, storing and analysing data. Additionally, more cost-effective data collection is a likely result from more efficient and better coordinated programmes. The establishment of regional sampling plans has, however, direct and indirect implications on the way data collection is currently organised between and within countries, and on the way data is stored, processed and disseminated to end-users. A key aim of the fishPi project was to support this transition.

An important part of the project was to undertake a regional consultation to compile views on possible future regional sampling plans as well as on different prerequisites, identified within fishPi, required to develop and implement such plans, determining the degree of consensus. Available expertise and resources required to develop and implement regional plans were further examined.

The consultation process was completed by written procedure through a questionnaire. The questionnaire was sent out to National Correspondents (NC) for data collection in MS and to heads of institutes from EU Member States fishing in ICES European waters (North Sea, North Atlantic and Baltic) as well as to ICES and the European Commission as they were considered main end-users of regional data. The overall response rate was 56%.

Overall, there was a high level of agreement in views expressed by NCs and heads of institutes. There was also strong support and consensus among the responding institutes and NCs on the general concepts of regional sampling plans based on statistical designs. Respondents agreed with the general approach towards probability-based sampling plans and probabilistic selection methods. Regional sampling plans are considered the best approach to improve quality of data and the cost-effectiveness of the current national sampling plans.

However, not everything can be regionalized. There was a strong consensus on the need to maintain elements addressing specific national requirements or objectives within the national plans even if the overall design is regional. Therefore, it is important to make sure that regional and national data collection plans and objectives are complementary.

There is a substantial amount of expertise, within key areas for development of regional plans, available at the institutes and MS. There is a willingness to make expertise available for development and implementation of regional plans but also some constraints, particularly where resources and available staff time are limited. Institutes and MS also have expertise to run and interpret scripts developed in the statistical environment R, as the ones developed within WP4 of fishPi.

Most respondents foresaw different problems related to the sampling of landings made by foreign vessels in national harbours. The main problem seems to be robust access to fish. Reasons for problems with access to fish are both administrative (permits to sample the fish) and logistical (fish rapidly transported to trucks). A second problem is access for the sampling country to relevant data contained in logbooks and sales slips. A third problem is that sampling of landings requires effective regional data management for samples to be processed and raised properly. In summary, sampling of foreign vessels requires more attention and dedicated actions.

There a strong consensus and support from the institutes and MS on the need for further development and implementation of a multi-purpose regional database (RDB). Some respondents see the RDB as a prerequisite for regional sampling plans. Significant cost saving associated with having a functional regional database where end-users can access data and remove the need for specific data calls/requests are also highlighted. Some respondents stress that national expertise need to be involved in the development and implementation of estimation procedures within the RDB. There is also a strong consensus on the need to develop the RDB further to include more quality assurance checks (as the ones developed within fishPi WP4).

Most respondents argue that additional resources will be needed to sample protected, endangered and threatened species (PETS) and/or stomach contents. The amount of additional resources is dependent on the level of ambition within the sampling plans. Priority levels need to be discussed and end-user needs have to be clearly defined. Respondents foresaw (or have experienced) some problems with access to vessels for PETS sampling. Foreseen problems are mostly related to potential refusals due to lack of space for observers on small vessels. Alternative sampling approaches might also be needed. All respondents answer that they have (or partly have, depending on the taxon) the relevant expertise for PETS identification. Some respondents highlight that training might be needed for observers. Expertise in stomach analysis is available in some institutes and MS. Resources to carry it out routinely and/or in large scale are however lacking.

1 Background and context

Member States have worked together since 2004 in Regional Coordination Meetings (RCMs) to coordinate national data collection programmes under the EU Data Collection Framework (DCF). In recent years it has become clear that regional data collection plans would offer a number of benefits over coordinated national plans. These benefits include more appropriate sampling designs, more rigorous statistical methodology, harmonised data formats and the adoption of standardised protocols. Regional fisheries sampling schemes would be more appropriate for providing data for the assessment of shared fish stocks in regional seas, they would provide a better understanding and greater transparency of the data collected, and they would allow for the proper evaluation of the quality of data and related estimates. More efficient and better coordinated data collection is also likely to be more cost effective.

The establishment of regional sampling plans has implications for the way data collection is organised between and within countries, and for the storage, processing and dissemination of regional data to end-users. The RCMs established under the DCF will evolve to Regional Coordination Groups (RCGs), as suggested in the recast of DCF, requiring more expert time and networking throughout the year to facilitate the development and management of regional programmes.

A key aim of the fishPi project is to support this transition. fishPi conducted simulation case studies to compare regional sampling designs for four regional fisheries, and these gave an insight into what may be required for the development of regional sampling designs for some of the shared stocks of commercial importance. The requirements of small scale and recreational fisheries data collection have been examined in relation to end-user requirements, as have collection schemes for data supporting ecosystem analysis. fishPi also further developed tools for standardised quality checks, data exchange formats and data harmonization. The outcomes of these various work packages identified some of the issues that need to be discussed and resolved before the implementation of regional sampling plans.

An important part of the project was to undertake a regional consultation to obtain views on possible future regional sampling plans as well as on different prerequisites, identified within fishPi, to develop and implement such plans with the aim to determine a degree of consensus. Available expertise and resources required to develop and implement regional plans were further examined. Outstanding issues that need to be addressed were noted.

The consultation process was completed by written procedure through a questionnaire. The main reasons for choosing this procedure were to ensure that as many voices as possible were being heard as well as to ensure sustainability of the results. The document could act as a basis for the evolving integration of regional sampling and the forging of cooperative regional links between Member States within a region. The results may also be of interest for national managers and end-users.

The questionnaire was sent out to National Correspondents (NC) for data collection in MS and to heads of institutes that carry out biological data collection identified from the EFARO (European Fisheries and Aquaculture Research Organisation) contact list. Heads of institutes and NCs from EU Member States fishing in ICES European waters (North Sea, North Atlantic and Baltic) were approached. The questionnaire was also sent to ICES itself and to the European Commission as they were considered the main end-users of regional data.

Overall, the questionnaire was sent to 15 National Correspondents. Answers were received from NCs in Estonia, Portugal, Denmark, Sweden, Finland, Ireland, Germany, Belgium, Netherlands, Poland and the UK resulting in a response rate of 73%.

Questionnaires were further sent out to 17 Head of Institutes with 8 answers received (IFREMER, SLU, MI, ILVO, IEO, AZTI, MSS and CEFAS) and a resulting response rate of 47%. However, some of the responses from the NCs were a joint response from the NC and the Head of the institute which means that this response rate in reality is higher.

Questionnaire was sent to 5 officials from the European Commission and to ICES. Responses were received from ICES and from one of the Commission officials.

The questionnaire comprised a set of questions on which respondents were asked to answer "yes", "no" or "partly". The respondents were also asked to provide relevant comments to the questions, in particular if they were answering "partly".

Responses to the questions in the questionnaire are summarized below. Specific comments frequently differed between respondents. In cases where two or more respondents made a similar comment, these were combined in the compilation below.

2 Questions and answers

1. The fishPi project proposes a move towards regional sampling plans which provides robust and transparent estimates to end-users using probability-based sampling schemes. A regional sampling plan includes: an integrated regional sampling design, standardised sampling protocols, a common approach to quality assurance, and regional tools for the management and dissemination of data. Benefits of this approach will include improved transparency and efficiency in the data collection schemes at a regional level.

A. Do you agree with the overall concept of a regional sampling plan as proposed above?

Support for the concept of regional sampling plans is strong. Most respondents answer "yes" (18/21) and a few "partly" (3/21). Comments made by respondents include:

- There are national objectives as well as regional within the sampling programme. Any regionalisation should be broad scale enough to accommodate these. This might be particularly true for stocks relevant to the MS but not subject to a European Total Allowable Catches (TAC). Minimum sampling requirements of each MS need to be contemplated and a regional sampling plan needs to be truly integrated with national requirements.
- Format/structure of regional work plans needs to be developed and details worked out. It is essential to develop these elements based on a common/compatible approach viable within each region because there is no "one size fits all" solution.
- Where possible, regional approaches should be integrated at EU-level.

There is a strong consensus among the respondents on the general concepts of regional plans. An area where there might be disagreement is how similar plans should be between regions (see last 2 bullet points).

B. From an end-user perspective, do you see benefits from the possible future establishment of regional plans?

Most respondents expect benefits from future establishment of regional plans. Most respondents answer "yes" (15/21), a few "partly" (3/21) and 3 give no answer (primarily as they do not consider themselves as end-users). Expected benefits include:

- The use of integrated regional sampling design, standardized sampling protocols, a common approach to quality assurance and regional tools for the management, processing and dissemination of data to have a significant positive impact on the quality of the data.
- We need to ensure that we learn from good examples of national plans.
- Regional plans to improve the possibility for ensuring collection of data from all types of fisheries in a standardized and transparent way.
- Data collection to become more aligned between MS within a region.
- Optimizing data collection between MS. Easier and more straightforward framework for adjustments and optimization.
- Regional plans to enable closer dialogue with end-users.
- More cost effective and efficient working and sharing of expertise.

ICES, as a main end-user of data in the North Sea, Atlantic and Baltic regions, responded to the question in the following way "As end-user the key aspect is not national versus regional plans but to have the data needed with the appropriate quality delivered on time. However, although it is not directly related to ICES, an important added value of the regional plan is the cost effectiveness, which will likely free resources to other activities, as for example data processing. The current resource's allocation to data processing is suboptimal and Member States have problems on delivering the data on time with the adequate quality evaluation."

There seems to be a substantial agreement on the expected benefits from regional plans. Key issues are transparency and improved quality of data, cost-effectiveness and optimization.

C. **From an end-user perspective, do you see risks from the possible future establishment of regional plans?**

Most respondents foresee potential risks from future establishment of regional plans. Most respondents answer "partly" (11/21), some "yes" (5/21), a few "no" (2/21) and 3 give no answer (primarily as they do not consider themselves as end-users). Comments cover a wide array of topics summarized in the bullet points below.

- Risk that data needed by national end-users (incl. national administrations) or other 'minor needs' are insufficiently covered under a regional plan. Respondents stress risk of loss of national control, e.g. when national landings are small in terms of overall contribution but important from national point of view. Small MS do not want to be ruled over.

- Risks that sampling levels for some stocks, fisheries, métiers or areas of interest might not meet the expectations of all end-users needs which could have an impact on specific advice. For a university or another type of end-user with a more local focus the specificities of a smaller scale could be lost.
- Risk that the regional plan is not fulfilling its objectives. A regional plan implies that every MS is dependent on sampling being adequately carried out in all MS involved, and this needs to be ensured. In terms of regionalisation of actually sampling plans, area or stock specific data collection might become completely dependent on a particular MS or laboratory with the associated risk of delivery linked to the management and available resources of a third party/partner.
- One of the effects of a fully implemented regional sampling frame would indicate re-distributing sampling effort across MS. Sampling is related with funds received, and it is reasonable that MS wants to ensure their funding. Sampling is also related with workload, and increased sampling responsibilities may not be affordable by a MS. When economical and technical arguments are mixed in the discussion, conflicts may arise.
- Different speeds in MS to adapt the data collection to the regional plans. MS within a region might have to change sampling schemes and procedures and it will take years before such changes can be expected to be fully implemented. Not all Member States might be able to implement changes at the same speed.
- Risk derived from the increased workload that this regional approach will imply. Would all MS be able to prioritise regional work within their institute to make expertise available for regional data collection plans?
- Communication between end-users and data collectors needs to improve greatly for the regional approach to succeed. This communication should be in both ways, and with all relevant end-users. This means that all relevant RMFOs need to be involved and their specificities (different needs, different speeds, etc.) taken into account in the regional approach.
- Some recognition needs to be taken of the administrative context within which this regionalization is conducted. Success will depend on how well the Commission, end users and national administrations and delivery bodies work together. The Commission should not be content to completely devolve responsibility entirely to regional groups. They will need to adjudicate in disputes, guide and compel cooperation from time to time and so should be more actively involved than they have been so far. If the regional plans are unsuccessful there is a risk that they become an additional bureaucratic layer where the regional plan becomes the focus, with the aim of adhering to specific design and protocols, at the cost of making sure that sufficient and useful data are collected to inform management and/or legislation.
- The Regional Coordination Groups have to adapt to their new roles.
- Risk that sampling plans cannot be handled efficiently on the data processing side as long as funds for development of RDB (Regional Database) is not in place.

Overall, Respondents foresee potential risks associated with the development and implementation of regional plans. Several respondents stress the need for regional plans to accommodate national objectives and requirements as well as regional. Another key-point is the perceived risk of different implementation speeds in different MS as MS become dependent on other MS. One respondent expressed that *“Probably, in the short term more than an adopted regional sampling plans it could be an on-going process at different speeds.”* A third key-point is re-allocation of sample obligations between MS as this has implications for utilization of funds. Finally, respondents highlight the need to explore the administrative side of regional plans as success will depend on how well MS, the Commission and end-users work together.

D. Do you think regional sampling plans are the best approach to improving the quality and cost effective collection of data required by end users? If not, what alternative approach would you propose?

Most respondents see regional sampling plans as the best approach to improve quality and cost-effectiveness in data collection. All respondents answer “yes” (15/21) or “partly” (6/21). Comments are summarized below:

- No respondent suggests alternative approaches but one respondent stresses that the approach should formally combine EU and regional levels. The regional approaches should be application cases of a global approach.
- Some respondents highlight the need to sample data to address national specificities as well as regional objectives. In the case of regional end users, regional sampling plans seem a good approach but a detailed analysis has to be carried out that checks of a regional approach can meet national requirements as well.
- A regional approach will most likely improve cost-effectiveness at least in the long term
- Overall quality improvement will depend on the schemes developed. One respondent expresses it as follows: “*the capacity of regional sampling plans to improve quality and provide good estimates will depend on a number of factors including adequate sampling designs, correct identification of the fishing activities, right sampling protocols, tough quality assurance, good implementation and appropriate raising procedures. Progress on most parts of these elements can be made based on the existing knowledge and regional capacities, thus providing an overall improvement. The approach to move forward the process for a regional sampling plan should take into account the different phases of the scientific knowledge around these factors: starting to implement those already well known and promoting the progress in the analysis of those where uncertainties persist. This process should ensure key steps, as the transition between all the nationals and the regional sampling design, are correctly addressed.”*
- To improve data quality respondents stress the need for resources to be allocated to setting up data screening routines and getting data into well-functioning regional databases.

- ICES has developed several guidelines to improve the quality of data collection. This work is ongoing and taking place under the remits of the ICES Planning Group on Data Needs for Assessment and Advice (PGDATA).

Overall, there is a general consensus among the respondents that regional plans are the best approach to improve quality and cost-effectiveness in data collection. Some respondents express that a regional plan *per se* does not increase quality.

2. An implication of the transition to regional sampling plans and standardised tools is that development work needs to be carried out at the regional level. This would represent a shift from the present situation where work is carried out at the national level. In the long run this is likely to ease the work load for the Member States as everyone could benefit from common developments. In the short term this will require human resources.

A. Can your country /laboratory dedicate e.g. 3 – 5 months of staff time with specific expertise to take part in a network to develop and implement regional sampling designs? If so please indicate what kind of expertise in the table below.

Most respondents are positive to dedicate staff time to the development and implementation of regional plans as they answer "yes" (6/21) or "partly" (8/21). One respondent answer "no" but commented that competence could be mobilised on a case by case basis within projects and processes. Six respondents did not give an answer, either as the question is not relevant for them (e.g., ICES, COM) or because they have not made a decision yet (depends on future EU-MAP and role of the RCGs). Comments are similar between the respondents regardless if the answer to the question is yes/partly/no or no answer and include:

- Willingness to prioritise regional work.
- Many respondents highlight limitations in resources and staff time. Participation has to be decided on case by case bases.
- Objectives and timely deliverables need to be clearly identified and well developed work plans are needed. Objectives of the regional plans need to meet both EU and national policy objectives, with a clear link between current staff resource allocation and the development of a regional approach.
- Means for efficient regional work need to be assured. In particular do we need to develop the regional database allowing for more efficient RCM work?

Respondents were asked on the available expertise within the MS on key aspects of a regional plan. They were further asked if they can make the expertise available for regional work. The answers are summarised in Table 1.

ICES highlights that several expert group meetings deal with quality criteria of the data collection and processing. ICES, also has a considerable network of experts that could contribute to the evaluations of the regional plans. Most of the areas listed in the table below are areas of priority listed in the ICES Strategic Plan 2014-2018.

Table 1. Showing a summary of available expertise within key areas for the development and implementation of regional plans. First answer relates to "Currently available in your institute (or institute(s) involved in the data collection". Second answer relates to "Are you willing to prioritise regional work within your institute to make expertise available for design and management of regional data collection plans?" Y = Yes, N = No and "-"= No response

Respondent	IT expertise i.e. database development and database management	Statistical expertise in sampling design and estimation.	Expertise in large scale commercial fisheries	Expertise in small scale commercial fisheries	Expertise in recreational fisheries surveys	Expertise in running sampling programmes	"R" programmers	Expertise in systematic quality assurance work	Expertise in surveying bycatch /PETS
Estonia (NC)	N / N	Y / N	Y / Y	Y / Y	Y / Y	Y / N	Y / N	N / N	Y / Y
Portugal (NC)	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -
France (IFREMER)									
Denmark (NC)	Y / Y	N / Y	Y / Y	N / Y	Y / Y	Y / Y	Y / Y	Y / Y	N / NA
Sweden (SLU)	Y	Y / Y	Y / Y	Y / Y	Y	Y / Y	Y / Y	Y / Y	Y / Y
Sweden (NC)	Y / Y	Y / Y	Y / Y	Y / Y	Y / -	Y / Y	Y / Y	Y / Y	Y / Y
Netherlands	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y
Finland (NC)	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y
Ireland (NC)	Y / Y	Y / Y	Y / Y	Y / -	Limited	Y / Y	Y / Y	Y / TBD	Y / TBD
EU									
Spain (IEO)	Y / N	Y / N	Y / N	Y / N	N / N	Y / N	Y / N	Y / N	N / N
Ireland (MI)	Y / N	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y
Germany (NC)	Y / N	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	N / N	Y / Y
Belgium (NC)	Y / Partly	Y / Y	Y / Y	N / -	Partly / Y	Y / Y	Y /Partly	Y / Y	N / -
Poland	Y / Y	Not yet / N	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y		Y / Y
ICES	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y	Y / Y
Belgium	Y / Y	N / N	Y / P	Y / P	Y / P	Y / P	Y / N	Y / Y	N / -
Spain (AZTI)	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -
UK (MSS)	Y / Y	Y / Y	Y / Y	Y / Y	N / N	Y / Y	Y / Y	Y / Y	Y / Y
UK (MMO)	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -
UK (CEFAS)	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -	Y / -

In conclusion: there is a large amount of expertise available at the institutes and MS. Most respondents are willing to make expertise available for development and implementation of regional plans but are sometimes constrained by limitations in resources and available staff time. Willingness is also dependent on objectives of the regional plans, means to effectively develop them and the future EU-MAP.

3. In a regional sampling design, the Regional Coordination Groups would identify an optimal allocation of national sampling commitments that would be most effective in meeting the needs for data to assess stocks and manage fisheries at a regional scale. Simulation studies within the fishPi project show that this might have an impact on how sampling effort is allocated between countries, and between different data collection activities within countries. For example this could include increasing or decreasing the overall number of sampling events, and/or changing the way the sampling activities are allocated.

   A. Is there flexibility (e.g. increase/decrease number of sampling events) in your country / laboratory to change the way in which your sampling is designed and implemented, to better meet the goals of a regional sampling plan?

   Most respondents answered that there is some flexibility to increase or decrease number of sampling events carried out by the institute or within the MS by answering "partly" (13/21) or "yes" (5/21). Comments primarily identify limitations to flexibility and include:

   - The need for sampling plans to meet regional but also national requirements and objectives. One respondent suggested that the regional sampling design should guarantee a minimal sampling level, but not override the requirements determined by the MS or institutes, since it cannot be guaranteed that the regional plan may deal with national (local) requirements.
   - Any sampling allocations will depend on the objective, sampling frame and units selected. A broad scale approach does not always capture national fleet segments and/or a regional management plan (which will be different for management areas). Improvements of quality and efficiency are paramount, but only if this is done at the appropriate scale to meet the fisheries management and policy objectives at different scales.
   - If a regional plan implies big changes due to, e.g., a high increase in sampling effort, it has implications for funding (not all institutes have flexibility to adjust sampling budgets) and the possibilities for the research institutes to assume the new requirements. It is essential to consider the latter because the National research institutes are involved in Data Collection issues but also have their own and more general objectives and limitations too.
   - Some respondents do not want to reduce their sampling effort.
   - The current national expertise is tailored to the present sampling design. New sampling designs might require changes in the national expertise, i.e. non-continuity of national expertise or the need to acquire new expertise (at what cost?).
   - The increase/decrease of sampling events should not be seen as <u>the</u> fundamental element of the regional approach. Progress in the regional approach can be made through common standards and procedures encompassing sampling designs, protocols, quality reviews, strengthening of the common formats and common tools (as InterCatch or the Regional Data Bases where they exist).

- EU Commission is hoping that MS laboratories will be able to show the necessary flexibility

In conclusion: most respondents seem willing to adapt their sampling schemes to take into account regional sampling plans, at least to a certain extent. There are however limitations. A key issue is the need to assure that sampling programmes meet national requirements and commitments as well. Another issue are budget constraints because not all respondents have flexibility in their sampling budgets to meet increased sampling commitments. Some respondents went further, in not wanting to decrease their level of sampling. One respondent highlighted that sampling allocation was just one part of a regional approach, and progress could be made on other aspects such as common standards and procedures encompassing sampling designs, protocols, quality reviews, strengthening of the common formats and common tools. This respondent was not comfortable with leaving all these decisions in the remit of the RCG.

4. The present national sampling plans are based on the current national obligations in the Data Collection Framework and may also serve end-user requirements at the national level.

A. Do you envisage that your national sampling plan would include elements that address specific national requirements as well as those addressing regional requirements?

Almost all respondents envisage that national sampling plans will include elements that address specific national requirements as they answer "yes" (18/21). One respondent answers "partly". ICES and COM did not respond to the question. Comments include:

- High importance for many institutes and MS.
- The same data collection framework must serve the regional and national needs together. One respondent stated that "Any good sampling scheme should maximise the output to serve as many objectives as possible. Servicing national policy objectives through a well-designed and structured sampling scheme is sensible as long it is not to the detriment of the original aims or objectives, or subsidised by the larger programme. It is about maximising data use and being aware of that while collecting primary data without increasing cost."

Overall, there is a strong consensus on the need for possibilities to include elements addressing specific national requirements within the national plans even if the overall design is regional. Preferably this should be done in a complementary and well-structured way.

5. The fishPi project suggests a move to statistically robust probability-based sampling plans which allow a thorough evaluation of data quality. This would involve the adoption of stratified sampling designs, and the use of sampling frames from which sampling units are selected using probabilistic selection methods.

A. Do you agree with the rationale behind this approach?

Most respondents answer "yes" (14/21) and some "partly" (6/21). One respondent did not answer. Comments include:

- Several institutes and MS have already implemented probability based sampling plans.
- Most respondents agree with the general concept, but foresee (or have experiences in) difficulties to implement these probability based sampling plans on the ground as the success will depend on external factors, e.g. access to

ports/vessels, costs of travel, facilities on vessels and in ports. The practical implications need to be considered.
- Each region should define the methodologies that better comply with its requirements. The approach might differ between regions. The best approach could be decided in the RCGs.

Overall, respondents agree with the general approach to move towards probability-based sampling plans and probabilistic selection methods. How these types of plan will work in practice needs, however, to be more thoroughly considered.

6. Probability-based selection requires random selection of sampling units, for example vessels at landing sites or from vessel lists.

A. Do you anticipate problems in adopting random selection protocols?

Most respondents answer "partly” (10/21) and some “yes” (5/21) or “no” (4/21). ICES and COM did not respond to the question. Comments were similar between respondent regardless of answer and include:

Comments:
- There are practical problems in implementing this such as, e.g., access to vessels, costs of travel, non-existing facilities onboard vessels, and refusal by skippers.
- Problems may occur for example in cases when the catch/landings to be sampled from selected vessel are already sold prior to landing and/or loaded directly and immediately from the vessel to the transport trucks.
- The challenge is not to adopt the protocols but to implement (and adjust) them in real life. This challenge shall not be underestimated in a regional context and might have implications for how regional random selection protocols can be adopted.
- Difficulties to control randomness in shore sampling programmes (selection of vessels in harbours and selection of boxes).
- Needs to consider local situations, and to preserve time-series continuity.
- Under achievements resulting in insufficient data for assessment purposes is not acceptable. Hence, priority needs to be given to main species/stocks/fisheries to ensure adequate data collection for management purposes.
- Introduction of landing obligation might increase problems.
- There are lots of associated “problems” with such methods but generally we have been able to find workable solutions.

Overall, respondents foresee problems associated with the adoption of random selection protocols. Several respondents have already implemented these types of protocols and have, at least to a certain extent, found solutions on the national level. Foreseen or experienced problems include access to vessels (willingness of fishermen to take observers) or catches at landing sites but also difficulties to control randomness within onshore sampling programmes. Some respondents highlight the risk of random selection protocols to be less time efficient.

B. Do you currently use a random selection procedure for selecting sites and/or fishing trips for on-shore sampling?

Most respondent answered “partly” (10/21) and some “yes” (5/21) or “no” (4/21). ICES and COM did not respond to the question. Comments include:

- Several respondents have implemented random selection procedures during the last years or are in the process of phasing in statistical sound sampling, at least for parts of their on-shore sampling programmes.
- Some respondents communicate that sampling design is based on using random selection procedures but sometimes it is not possible to adhere to it due to practical issues.

In conclusion, most institutes and MS are well on their way towards implementation of random selection procedures in on-shore sampling programmes. Some respondents express that they sometimes have difficulties to sample in a random way in the field. However, a few institutes or MS have not yet implemented random selection procedures in their shore sampling programs.

C. Do you currently use a random selection procedure for selecting vessels for at-sea sampling?

Most respondents answered “yes” (9/21) or “partly” (5/21). A few respondents answered “no” (3/21) or did not respond (4/21, including ICES and COM). Comments include:

- Most respondents report that they have implemented random selection procedures in their sea-sampling programme for the entire or parts of the fleet.
- Some respondents report that the selection procedure is pseudo-random due to different logistical constraints.
- Some institutes and MS have not implemented random selection procedures.

In conclusion, most institutes and MS are well on their way towards implementation of random selection procedures in sea-sampling programmes. Some respondents select vessels to sample in a pseudo-random or opportunistic way. A few institutes or MS have not yet implemented random selection procedures within sea-sampling programs.

7. Information on non-response and refusals needs to be collected in order to assess the overall performance of a probability-based sampling program. An example of non-response is where a vessel is selected to take an observer but cannot be contacted. A refusal would be where the skipper is contacted but refuses to take an observer.

A. Are you already calculating non-response and refusal rates?
Most respondents answered “yes” (9/21) or “partly” (6/21). A few respondents answered “no” (4/21). ICES and COM did not respond to the question. Comments include:

- Most respondents report that they calculate non-responses and refusals. Some respondents do it in some areas and/or for some part of their sampling programme.
- There is a need to have standard protocols to enable comparable estimates across the region.

- A few respondents report that they do not calculate non-responses and refusals.

In conclusion, most institutes and MS are calculating non-responses and refusals, at least to some extent. One respondent highlighted the need for standard protocols enabling comparisons between institutes and MS within a region. A few institutes or MS, primarily in the Baltic region, do not calculate non-responses and refusals.

B. If you are not collecting non-response and refusal data, do you think it will be possible to collect this information?

Most respondents did not answer the question (12/21) as they presently are collecting non-response and refusal data. The other respondents answered “yes” (8/21) or “partly” (1/21). Comments include:

- Very time consuming
- Problems foreseen in case of small scale fisheries

In conclusion institutes and MS that presently are not collecting data on non-responses and refusals find it will be possible to do so.

8. Regional sampling designs are likely to assume that individual countries will include the landings of all flag vessels into their country in their on-shore sampling plan.

A. Do you envisage any problems with sampling the landings of foreign vessels?

Respondents answered “partly” (11/21), “no” (4/21) and “yes” (3/21). Three respondents (including ICES and COM) did not answer. Comments include:

- Normally these landings are not processed in the same way as national landings. When and if the landings are processed or presented for sale in the same way as national landings then they could be included in the same selection process.
- Access to the fish for sampling (i.e., bureaucracy and permits); on-shore sampling tends to be on a voluntary basis from the point of view of the fishermen. Access to foreign vessels needs to be administratively and legally regulated to avoid refusals from foreign flag vessels.
- Access to the fish for sampling; landings are often going straight in the back of a lorry to be transported back to the flag country for sale.
- No problems when the flag vessels landed in the harbour. Problems occur when the landing arrives by trucks from other countries. In these cases sometimes trips and vessels are very difficult to identify because trips and different vessels landings are mixed together.
- Tracking of landings transported by trucks is essential. A system identifying the trips that the different landings (logbooks) and sales (sales notes) belong to, would be very useful.
- Some organisational change will be needed to ensure that all Marine Labs have direct access to the appropriate transversal (logbooks and sales slips) data. Care needs to be taken at this point that such an approach is supported by legislation

- A difficulty occurs when landings are not complete trips but partial. Landings can be sold partially in different countries and harbours.
- It is not known in advance if, when and where such landings may take place.
- Sampling of foreign vessels requires an effective regional data management (implemented regional database).
- Limited sampling resources of a MS could make complying with regional sampling designs not always possible.

In conclusion, most respondents foresee different problems related to sampling of foreign vessels. The main problem seems to be robust access to fish. Reasons for problems with access to fish are both administrative (permits to sample the fish) and logistical (fish rapidly transported to trucks or processed in different ways compared to national landings). Rapid transport of fish from the landing site for sale in other countries (e.g., flag country) are highlighted as an issue by several respondents. When landings are processed and presented for sale in the same way as in the landing country it seems the actual sampling does not constitute a major problem. A second problem is access for the sampling country to relevant information in logbooks and sales slips. A third problem is that sampling of foreign landings requires effective regional data management for samples to be processed and raised properly. A fourth problem is limited sampling resources within the MS. In summary, issues related to the sampling of foreign vessels need more attention, in particular if they are to be part of integrated regional sampling plans. Care needs to be taken with regard to the legal aspects (e.g., access to transversal data) as well. Dedicated procedures might need to be developed.

9. The sampling of commercial marine fisheries at a regional level, using stratified designs and employing probabilistic selection methods, will require the adoption of common protocols for the selection of sampling units and the collection of biological information. Such protocols should contain specific details and references to enable anybody to find all the information required to correctly collect the data required.

A. Do you envisage a problem with moving towards regionally defined data collection protocols?

Respondents answered “no” (13/21), “yes” (3/21) and “partly” (2/21). Three respondents (including ICES and COM) did not answer. Comments were similar between respondents regardless of answer and include:

- The caveat is that regional protocols have to be implementable at national level, if they are too prescriptive and inappropriate it will not work.
- It will only be completely adopted if it does not jeopardise data collected in support of national or regional policy objectives. It is important to make sure that data collection programmes and objectives are complementary; otherwise the process will be counterproductive with increased cost and decreased efficiency.
- The regionally defined data collection protocols have to improve current MS protocols.

In conclusion, respondents are positive to regionally defined data collection protocols as long as they are flexible enough to enable MS and institutes to sample for its own needs (for example comply with national end users requirements). They also need to improve the present national protocols. It is important to make sure that regional and national data collection programmes and objectives are complementary.

10. The development of regional sampling plans as proposed by the fishPi project involves the simulation testing of sampling designs prior to their implication in practice. This requires the short-term sharing of anonymised trip-level logbook and sales data and sampling data solely for this purpose. In the fishPi project this was achieved through a carefully worded data-sharing agreement specifying the use of the data, the timeframe over which it could be used, and a password-protected website.

    A. Do you envisage a problem of the short-term sharing of trip-level and sampling data if appropriate data-sharing and security agreements and are put in place?

    Respondents answered “no” (12/21), “partly” (5/21) and “yes” (2/21). COM and one institute did not answer. Comments include:

    - Involved in the fishPi project and this system worked well.
    - For several respondents the short term sharing of trip-level data is dependent on authorization by national administrations.
    - Possible as long as confidentiality is ensured and that there are appropriate safeguards to ensure information cannot be used to identify people or/legal entities.
    - Provision of trip level transversal data goes beyond what is envisaged in the DCF legislation. No problem in principle in supplying data at this level but protocols for data protection would need to go some way beyond a data sharing agreement and password protected website. Protocols already exist under the European Statistical System. There is also a raft of relevant EU legislation in this area that should be referred to.
    - No problem to share biological information.

    ICES comments that it has experience with obtaining some of these types of data via its Data calls in connection with fisheries and environmental management questions posed by clients of advice. With appropriate measures it agrees that there should not be a problem with the short term sharing of these data.

    Overall, several of the respondents that answered that they do not envisage problems with short-term sharing of trip level data (provided proper data-sharing and security agreements in place) have not provided comments. These respondents were also partners in the fishPi project and it is interpreted that the system used in fishPi has worked satisfactory. Short-term sharing of trip level data is for some respondents dependent on national authorities which might create problems in some cases. One respondent highlights that provision of trip level transversal data goes beyond the envisaged in the DCF legislation. Protocols for data protection might need to go some way beyond a data sharing agreement and password protected website. Protocols already exist under the European Statistical System.

11. Regional databases (RDBs), in which all regional fisheries dependent data are stored in common formats are presently used for coordination purposes by the Regional Coordination Meetings. The RDB has considerable potential to support Member States in evaluating the

quality of data, and the generation of estimates from biological data needed by end users. This would allow for consistent and transparent responses to data calls and other types of reporting.

A. Do you support the idea of providing data to the RDB for multiple purposes linked to the DCF requirements (regional coordination and cooperation, quality assurance, production of annual reports, production of estimates for stock assessment)?

Respondents answered “yes” (17/21) or “partly” (4/21). Comments include:

- A RDB is a prerequisite for regional coordination and cooperation, clear difficulties to see either the implementation of a regional plan or the benefits of such a plan without the use of a multipurpose RDB.
- This is of greater benefit to data collectors and end-users than allocating resources to harmonize detailed sampling designs and generating generic protocols. Significant cost saving associated with having a functional regional database where end-users can access data and remove the need for specific data calls/requests.
- This must be done through agreements preserving the relevant commitments.
- Considerable MS specific expert input is likely to be a requirement to produce estimates for stock assessments. Blind assessments without regard to the implications of data subsets are ill-advised. Estimates are national responsibilities.
- Subject to seeking approval from data owners/relevant experts in each case that the data are fit for the purpose they are being used for.
- Important not to forget that a RDB has been agreed and is in use only in the North Atlantic, North Sea and the Baltic. To date, no RDB has been agreed for the Mediterranean or other areas/RFMOs.

ICES fully supports an RBD with multiple purposes and made the following comment “ICES hosts the Regional DataBase, RDB, which contain EU data collected from Baltic, North Sea and North Atlantic. The ICES Council agreed to fund further development of the regional database in 2016. The developments focus on harmonisation and standardisation of code lists, solving important issues and considering new catch categories resulting from landing obligations.”

COM makes the following comment; It could be a good idea to have a regional database holding primary sampling data which is assembled on a confidential basis in order to implement cross-MS interpolation procedures, and derive interpolated values for various parameters for non-sampled or under-sampled units. Such a database could be used to create databases of interpolated or fitted data which could be returned to each Member State. The regional database of primary data would then be destroyed (though the interpolation or fitting procedure would be documented). Member States would then be able to supply to end-users, on request, aggregated data which includes the interpolated values to replace missing observations. This could be done for any requested aggregation level, with no manual intervention on the data, and still producing consistent results across all aggregations requested. If necessary, regional databases holding data at an agreed level of aggregation could also be maintained for certain standard assessment tasks.

In conclusion, there is a strong support from the respondents on the implementation of a multi-purpose regional database. Some respondents see the RDB as a prerequisite for regional sampling plans. Significant cost saving associated with having a functional regional database where end-users can access data and remove the need for specific data calls/requests are also highlighted. A concern expressed by some respondents is the need for input from national experts when preparing estimates for stock assessment as local knowledge is required (and estimates responsibility of individual labs). Use of the RDB for regional cooperation, quality assurance and production of annual reports is not considered problematic.

12. The use of probability-based sampling schemes will require the additional collection and provision of information not currently required to be reported as part of the DCF, for example information on sampling events, sampling probabilities, the strata to which sampling units belong, and stratum sizes. This information would have to be included in national data bases so that it can be transferred to the regional data base.

A. Do you envisage a problem in adapting your national database to the requirements of a probability-based sampling scheme?

Respondents answered “partly” (8/21), “no” (7/21) or “yes” (4/21). COM and ICES did not answer. Comments include:

- Changes to national database are extremely time consuming and complicated.
- Overall, the adaptations should be technically feasible. The requirements might be considered case by case to define relevant supplementary tasks and associated costs.
- Adapting the national database requires time and funding. A thorough sampling scheme needs to be established and agreed upon prior major changes are made to the databases.
- Some respondents’ databases have recently been or are presently being updated to enable recording of information needed for probability based sampling (at least in ICES regions).
- There are difficulties with the integration of different data sources that are used in the decision making process as part of the sampling design. E.g. when the sampling design is based on log book data, linking sampling trips with log book data can be problematic.
- The incorporation of “near real time data” can also be problematic.

The extent of envisaged problems adapting national databases to requirements of probability-based sampling schemes, differ between respondents. Some respondents have already adapted their databases or are on the way to adapt them while others have not. Adaptation requires time and funding. A clear view on the requirements is needed before some respondents are willing to adapt their databases.

13. The move to a regional sampling design allows for quality assurance at a regional level, using consistent methods applied across national data sets. In particular it would allow estimates of the age and length composition of catches, and the quantities discarded, to be calculated at a regional level using estimators appropriate to the sampling protocols employed. At present

national estimates are derived independently which makes objective assessment of the appropriateness of such methods difficult.

A. Do you support the development of the current RDB to include regional quality assurance checks?

Respondents are positive to develop the current RDB to include regional quality assurance checks as they answer "yes" (19/21) or "partly" (2/21). Comments include:

- Initiative fully supported, it is long overdue. The management of regional fisheries has to be based on the use of correct estimates and quality assured data.
- Regional data quality may be useful but national expertise has to be included.
- There may be no benefits for local (G2/G3) stocks.

ICES fully support the concept of an RDB that includes the necessary detail data and tools for an adequate data quality evaluation. The RDB hosted by ICES allows several quality checks but more and regional checks would increase the overall data quality.

In conclusion, there is strong consensus among the respondents on the need to develop the RDB to include regional quality assurance checks. National expertise should be included in the development.

B. Do you support the development of the current RDB to include the appropriate statistical estimation at a regional level?

Respondents are positive to the development of the current RDB to include appropriate statistical estimation procedures at regional level as they answer "yes" (15/21) or "partly" (5/21). COM did not answer the question. Comments include:

- This should be considered as application case from procedures defined at the global EU level.
- It is a prerequisite for ensuring transparency and standardisation. If further development with the RDB is to take place this will need extra funding. However, the impression that this will be a cost effective method in the long run.
- Especially for G1 species.
- Need to have complete confidence in the methodology used, and do not see that this can be done without considerable input of national experts into that process.
- Considerable MS specific expert input is likely to be a requirement to produce estimates to stock assessments. Institutes involved in sampling are the ones who better know the fisheries they are targeting, and potential implementation problems encountered.

ICES fully support the concept of an RDB that includes the necessary detail data and tools for an adequate data quality evaluation. The RDB hosted by ICES allows several quality checks but more and regional checks would increase the overall data quality

In conclusion, there is consensus among the respondents to develop the RDB to include appropriate statistical estimation procedures at the regional level. However, respondents also highlight that they need to have complete confidence in the methodology used, and do not see that this can be done without considerable input of national experts into that process. National expert input is likely to be a requirement to produce estimates to stock assessments as institutes involved in sampling are the ones who better know the fisheries they are targeting as well as potential implementation problems encountered.

14. fishPi has developed scripts based on the statistical environment "R" to be used for quality checking and validating sampled national and regional data.

   A. Do you currently have the expertise in "R" to run quality checks developed in fishPi and to interpret the outputs?

   Respondents do have expertise to run and interpret scripts developed in "R" as they answered "yes" (18/21) or "partly" (2/21). COM did not answer. Comments include:

   - Several institutes have developed competences or large competences "R" and use it commonly.
   - Experts able to work with such programs but we do not have practical experience with fishPi, so we would apply "precautionary approach" here.
   - Regular training is also recommended in order to ensure that expertise is improved and maintained in the institutes.
   - Routine controls should be implementable through user-friendly interfaces.

   ICES has the expertise in house and in its network to utilize the scripts developed by fishPi. While a workable script, in whatever language, is important at the implementation level, ICES would stress that adequate documentation of the quality checking and validation is equally important to the end user. ICES see the possibility to run R scripts as an important part of the continuing development of the methods in the RDB. The aim is to include approved R scripts into the RDB, so the methods used are standardized, approved and version controlled. The methods developed should be documented and tested and finally approved by a group of statistical methods experts, e.g. WGCATCH before implemented in the RDB.

   In conclusion, all responding institutes and MS have expertise to run and interpret scripts developed in the statistical environment R. However, respondents that were not partners in fishPi have difficulties to assess if they can interpret the scripts developed within this specific project. ICES stress that proper documentation of the quality checking and validation routines is important to the end user. ICES see the possibility to run R scripts as an important part of the continuing development of the methods in the RDB. The methods developed should be documented and tested and finally approved by a group of statistical methods experts, e.g. WGCATCH before implemented in the RDB.

15. A stronger emphasis on ecosystem data, in particular, data on by-catches of protected, endangered and threatened species (PETS) and data on stomach contents, is foreseen in the future data collection regulation. The fishPi project has been working to identify end-user needs for this data and has been elaborating on potential future sampling plans.

A. Do you think that sampling for PETS and stomachs can be conducted with your available resources or will it require additional resources?

Respondents answered “partly” (10/21), “no” (6/21) or “yes” (3/21). Almost all respondents comment that they foresee needs for additional resources for sampling of stomachs and/or PETS. The answer is most likely dependent on if the respondent answered the first or second part of the question. COM and ICES did not answer. Comments include:

- Several respondents argue additional resources needed to sample PETS and/or stomachs.
- Dependent on whether PETS information is collected as an integrated part of the data collection for the CFP. Several respondents are already sampling PETS within present at-sea sampling programme. If special data collection schemes for PETS are needed in addition to the CFP data collection, they will require additional resources.
- Stomach sampling and analysis is presently not conducted in several of the responding institutes or MS and will require additional resources.
- Stomach analysis involves far more resources than sampling.
- If stomach sampling is at a limited level (limited number of species and it is not carried out annually) it can be carried out within the available resources. Any larger stomach sampling programme needs additional funding.
- Strict, pre-defined end-user needs and use (!) are a pre-requisite for starting stomach collection and analysis.
- It will be necessary in a first stage to confirm the priority level of PETS and stomach contents sampling.

ICES support this initiative which among others supports the ICES Working Group on Bycatch (WGBYC). Moves to integrate monitoring of the bycatch of protected species in all EU waters within the Data Collection Framework will require significant adjustments from those used to monitor commercial fish bycatch. Specifically-designed monitoring schemes (including dedicated observers or Remote Electronic Monitoring) are required if good estimates of protected species bycatch are required.

In conclusion, most respondents agree that additional resources are needed for sampling (and analysis) of PETS and stomachs. The amount of additional resources is dependent on the level of ambition within the sampling programmes. Priority levels need to be discussed and end-user needs clearly defined.

B. Is there appropriate access to fishing vessels to sample by-catch of PETs species? (The bycatch of these species can be a sensitive issue and refusal to collect data on these species on commercial vessels is quite common).

Respondents answered “partly” (12/21), “yes” (5/21) or “no” (1/21). Three respondents, including COM and ICES did not answer. Comments include:

- There are problems to sample, using observers on-board in some métiers/fisheries.
- We have observers on board some fisheries where it is mandatory to collect information on by-catch of PETS and also CCTV installed. As it is mandatory and beneficial for them, in these cases we don´t have important problems.

- The introduction of PETS data in the new DC MAP, will imply new fisheries to sample and in some cases specific monitoring for PETS that would probably result in difficulties to go on-board and increase refusal rate.
- There is some 'rumour' of refusals as a consequence of PETS but it is not wide spread.
- Vessel access is good at this point in time, but in general expected to change across the EU fleet with the introduction of the landings obligation.
- Sampling design and stratification has been designed for sampling commercial species and it does not allow the adequate quantification/sampling of the bycatch of PET species. Passive gears are prime candidates for bycatches of PETS, not presently included in some respondent's sea-sampling programme.
- In many cases by-catch of PETs species occurs in the small scale fishery fishing close to shore. This might imply small vessels need to be sampled and risk of lack of space for observers. Alternative ways to do PETS sampling might be needed.

In conclusion respondents foresee (or have experience of) some problems with access to vessels for PETS sampling. Foreseen problems are not only related to potential refusals but also, in several cases, associated with the fact that bycatches of PETS are believed to occur in parts of the fleet which are presently not sampled extensively in the sea-sampling programmes. A particular issue might be lack of space for observers on small vessels. Alternative sampling approaches might be needed.

C. Is there stomach content analysis expertise and resources available in your institution?

Respondents answered "partly" (11/21), "yes" (4/21) or "no" (3/21). Three respondents, including COM and ICES did not answer. Comments include:

- Expertise available but the resources are limited and is not available for all fisheries/areas.
- There are experts but if the collection of this data means a big increase in the analysis, more experts will be needed. Nowadays these experts work for specific projects but not as a routine data collection programme on stomach content.
- Dependent on trophic level of analysis (e.g. stomach in stomach) but in general the expertise is available.
- Some expertise, even though stomachs are not analysed on a routine basis.
- With regional training workshops the relevant expertise could be established reasonably quickly.

ICES host a database on stomach content.

In conclusion, there is expertise available in some institutes and MS. However, resources are lacking as stomach analysis is not part of the present regular sampling programmes. Several respondents express that they have limited expertise so stomach analysis is not done on a routine basis. Training workshops might improve the situation.

D. Is there PETS identification expertise available in your institution?

Respondents answered “yes” (11/21) or “partly” (8/21). COM and ICES did not answer the question. Comments include:

- Expertise available but the resources are limited.
- Bycatches are recorded according to supplied species lists, though the emphasis is generally on marine mammals, larger fish, and odd things. It would be fair to say that our at-sea observers are less adept at recording seabirds.
- For some taxa.
- Some training may be needed for observers.

In conclusion, all respondents answer that they have or partly have the relevant expertise for PETS identification. However, in some cases resources are limited. Some respondents indicate that might be a difference between taxa. Some respondents highlight that training might be needed for observers.

3 Summary of results

All results are summarized in Figure 1.

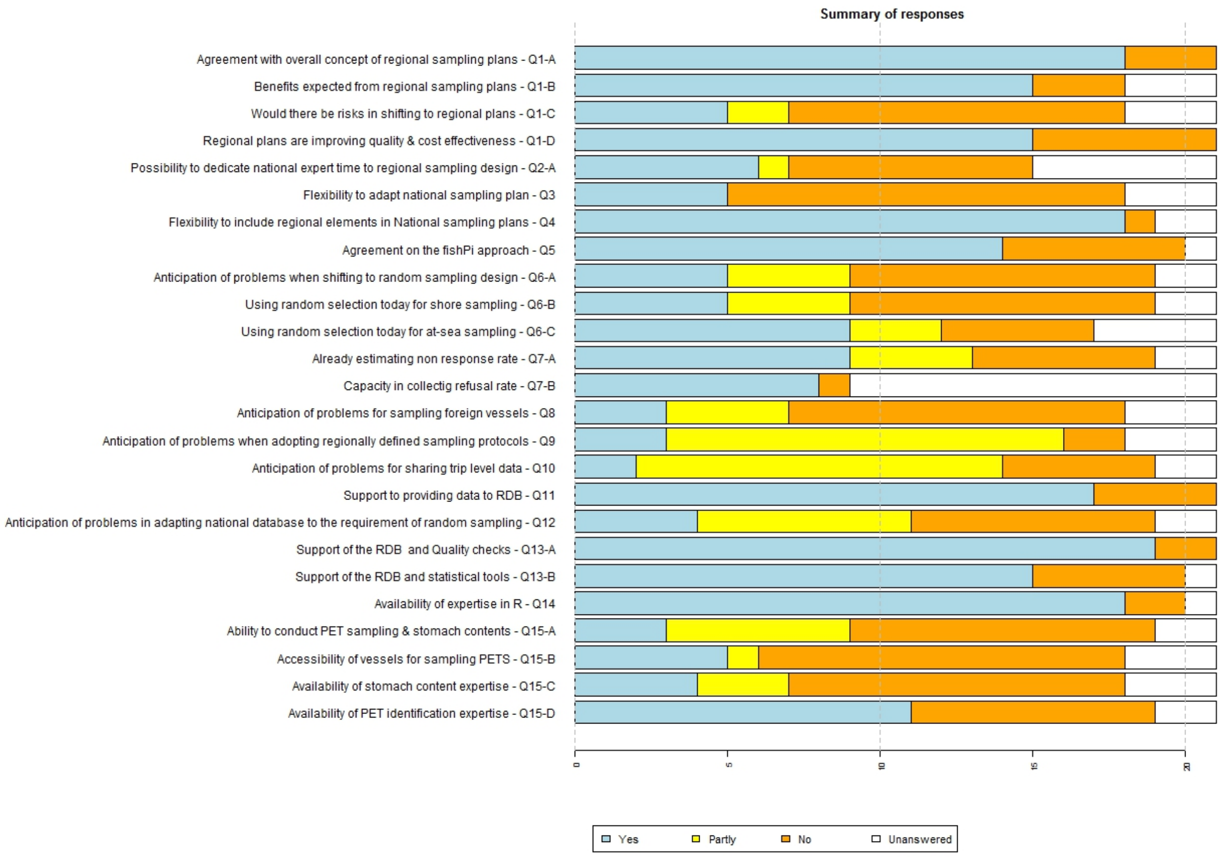


Figure 1. Showing a summary of responses on the fishPi questionnaire on views and key-aspects of regional sampling plans. Blue =”yes”, yellow=”no”, orange=”partly” and white = no answer. Note that for some questions (6A, 8, 9, 10, 12) are “no” a positive answer. For full questions see questionnaire.

A positiveness index was constructed to get a better overview on the level of agreement as some questions were formulated " do you agree with… (yes = a positive answer) and others "do you see a problem with… (no = a positive answer). In the positiveness index we count the positive answers. A "yes" (or "no" in the cases were no is a positive answer) got 1 score points while a partly get 0,5. Non - answers are deducted from the analysis. Results are then normalized to the maximum possible amount of score points given the number of answers for each question. This implies that the maximum score a question can get is 1. The positiveness index is presented in figure 2. The positiveness index varied between 0.93 and 0.45 for different questions. Out of the total of 25 questions, 10 got a higher score than 0.8 while 3 got a lower score than 0.5. Questions that got high scores (and respondents replied positively to) were for regional sampling plans to include national requirements as well as regional (Q4), support for the RDB (Q11, Q13A, Q13B), agreement of overall concept of regional sampling plans (Q1A) and availability within institutes and MS of expertise in "R" (Q14). Questions that got low score (issues that respondents are more concerned about) were risks associated to the shift towards regional sampling plans (Q1C), ability to perform PETS and stomach content sampling without additional resources (Q15A) and adoption of random selection protocols (Q6A). The postitiveness index is presented in figure 2 and a ranking of positive replies to the different questions in figure 3.

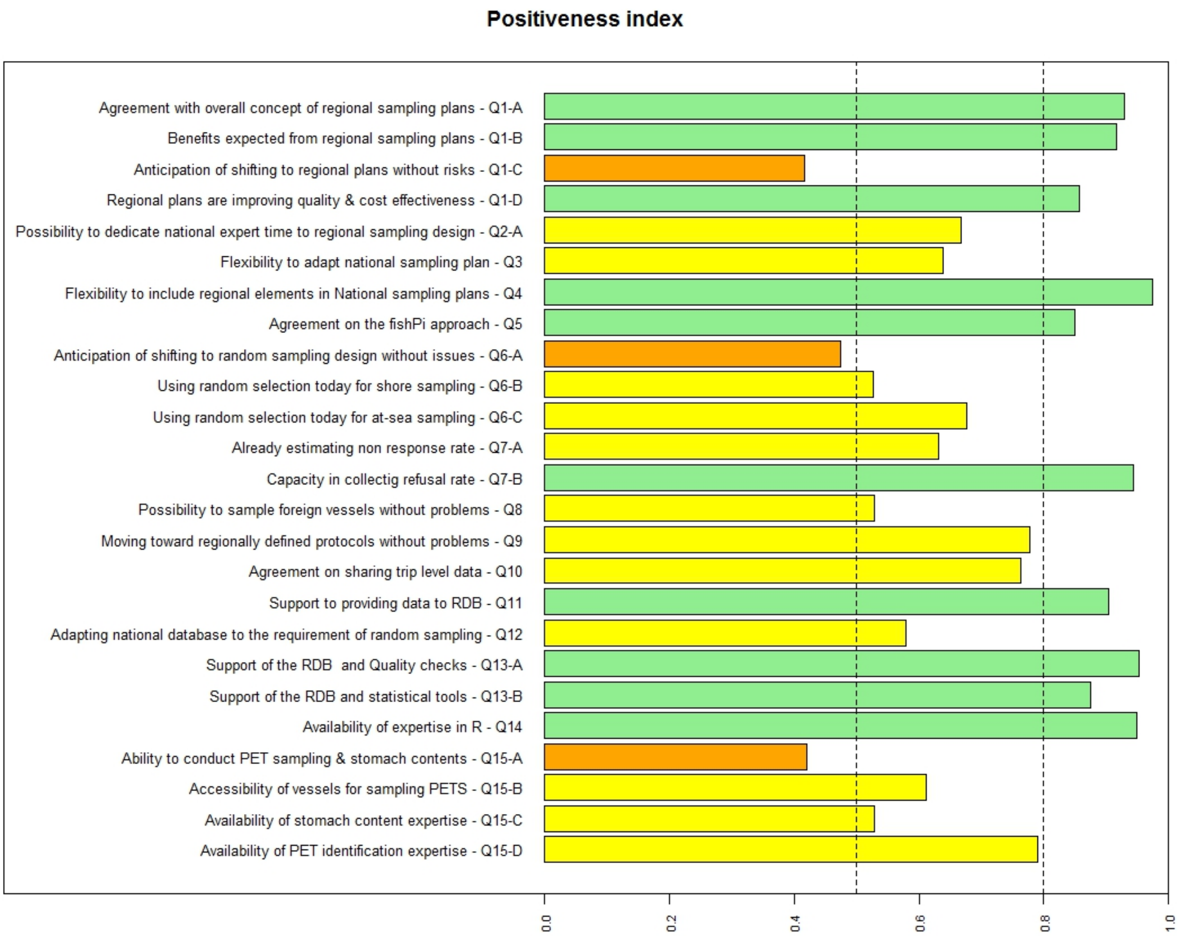


Figure 2. Showing positive replies for the fishPi questions in the fishPi questionnaire. The maximum score a question can get is 1.

Figure 3. Showing positive replies for the fishPi questions in the fishPi questionnaire ordered by the degree of positiveness. The maximum score a question can get is 1.

4 General conclusions

There is a strong support and consensus among the responding institutes and MS on the general concepts of regional sampling plans. Such plans are considered the best approach to improve quality of data and the cost-effectiveness of current national sampling plans.

An area where there appears to be some disagreement is around how much integration plans would require between regions. One respondent argued that regional approaches should be integrated at EU-level where possible while another found it essential to develop elements of a plan viable within each region because “one size does not fit all”.

There is a strong consensus on the need to include elements addressing specific national requirements or objectives within the national plans even if the overall design is regional. It is important to make sure that regional and national data collection plans and objectives are complementary. This is a pre-requisite for several respondents.

Respondents do foresee potential risks associated to the development and implementation of regional plans. One perceived risk is different implementation speeds in different MS with MS sampling becoming dependant on sampling carried out by other MS. Another key-point is reallocation of sample obligations between MS as this has implications for utilization of funds. Finally, respondents highlighted the need to explore the administrative side of regional plans as success will depend on how well MS, the Commission and end-users work together. Risks identified need to be addressed satisfactorily in work-plans and during the development phase.

Respondents are, to some extent, willing to adapt their sampling commitments to meet regional objectives in regional plans. There are however limitations. A key issue is the need that sampling plans meet national requirements as well. Another issue is budget constraints, not all respondents have flexibility in their budgets to meet increased sampling commitments. Some respondents do not want to decrease their level of sampling. One respondent is not comfortable with leaving all decisions on sampling commitments to the RCG.

There is a substantial amount of expertise, within key areas for development of regional plans, available at the institutes and MS. There is a willingness to make expertise available for development and implementation of regional plans but sometimes constraints such as resources and available staff time is limited.

Respondents agree with the general approach to move towards probability-based sampling plans and probabilistic selection methods. How these types of plans will work in practice needs however be more thoroughly considered. Foreseen or experienced problems include access to vessels (willingness of fishermen to take observers) or catches at landing sites but also difficulties to control randomness within onshore sampling programmes. Some respondents highlight the risk of random selection protocols to be less time efficient. Most respondents have already, at least partly, implemented random selection procedures. Some respondents sample in a pseudo-random or opportunistic way. Most institutes and MS are calculating non-responses and refusals, at least to some extent. One respondent highlighted the need for standard protocols enabling comparisons between institutes and MS within a region. A few institutes or MS, do not calculate non-responses and refusals but find it possible to do so.

Most respondents foresee different problems related to the sampling of foreign vessels. The main problem seems to be robust access to fish. Reasons for problems with access to fish are both administrative (permits to sample the fish) and logistical (fish rapidly transported to trucks). A second problem is access for the sampling country to relevant information in logbooks and sales slips. A third problem is that sampling of foreign landings requires effective regional data management for samples to be processed and raised properly. In summary, issues related to sampling of foreign vessels need more attention and dedicated actions.

Most respondents do not envisage problems in the short-term sharing of anonymised trip level data (provided proper data-sharing and security agreements in place) needed for simulation and design of regional sampling plans. This was completed within fishPi and the system seems to have worked satisfactory. Short-term sharing of trip level data is for some respondents dependent on national authorities which might create problems in some cases. One respondent highlighted that provision of trip level transversal data goes beyond the envisaged in the DCF legislation.

There is a strong consensus and support from the institutes and MS on the need for further development and implementation of a multi-purpose regional database. Some respondents see the RDB as a prerequisite for regional sampling plans. Significant cost saving associated with having a functional regional database where end-users can access data and remove the need for specific data calls/requests are also highlighted. Some respondents stress that national expertise needs to be included in the development and implementation of the RDB. This is particularly true for development of the RDB to include appropriate statistical estimation procedures at the regional level (estimates for assessment). Respondents highlight that they need to have confidence in the methodology used, and do not see that this can be done without considerable input of national experts. There is also a strong consensus on the need to develop the RDB further to include more quality assurance checks and for the production of annual reports.

Institutes and MS were asked to what extent they envisage problems to adapt national databases to requirements of probability-based sampling schemes. Answers differ between respondents. Some respondents have already adapted their databases or are on the way to adapt them but others have not. Adaptation requires time and funding. A clear view on the requirements is needed before some respondents are willing to adapt their databases.

Institutes and MS were asked if they had expertise to run and interpret scripts developed in the statistical environment R, like the ones developed within WP4 of fishPi. In conclusion all responding institutes and MS do have this type of expertise.

A stronger emphasis on ecosystem data, in particular, data collection on by-catches of protected, endangered and threatened species (PETS) and data on stomach contents, is foreseen in the future data collection regulation. Such emphasis will imply new sampling commitments and new types of expertise. Most respondents argue that additional resources will be needed to carry out these new types of sampling. The amounts of additional resources are dependent on the level of ambition within the sampling plans. Priority levels need to be discussed and end-user needs clearly defined. Respondents foresee (or have experienced) some problems with access to vessels for PETS sampling. Foreseen problems are not only related to potential refusals but also, in several cases, associated with the fact that bycatches of PETS are believed to occur in parts of the fleet which presently are not sampled extensively in sea-sampling programmes. A particular issue might be lack of space for observers on small vessels. Alternative sampling approaches might be needed. All respondents answer that they have or partly have (dependent on taxa) the relevant expertise for PETS identification. Some respondents highlight that training might be needed for observers. Expertise in stomach analysis is available in some institutes and MS. Resources are however lacking as stomach analysis is not part of the present regular sampling programmes. Training workshops might improve the situation.

ICES is a key end-user that will support several aspects of regional sampling plans. ICES has developed several guidelines to improve the quality of data collection. This work is ongoing and taking place under the remits of the ICES Planning Group on Data Needs for Assessment and Advice (PGDATA). ICES has several expert group meetings dealing with quality criteria of the data collection and processing. ICES, also, have a considerable network of experts that could contribute to the evaluations of the regional plans. Most of the areas listed in the expertise table are also areas of priority listed in the ICES Strategic Plan 2014-2018.

ICES fully supports an RBD with multiple purposes. ICES hosts the Regional DataBase, RDB, which contains EU data collected from Baltic, North Sea and North Atlantic. The ICES Council agreed to fund further development of the regional database in 2016. The developments focus on harmonisation and standardisation of code lists, solving important issues and considering new catch categories resulting from landing obligations. ICES fully supports the concept of an RDB that includes the necessary detail data and tools for an adequate data quality evaluation. The RDB already includes several quality checks but more (and more regional) checks would increase the overall data quality. ICES has the expertise in house and in its network to utilise the scripts developed by fishPi. While a workable script, in whatever language, is important at the implementation level, ICES would stress that adequate documentation of the quality checking and validation is equally important to the end user. ICES see the possibility to run R scripts as an important part of the continuing development of the methods in the RDB. The aim is to include approved R scripts into the RDB, so the methods used are standardised, approved and version controlled. The methods developed should be documented and tested and finally approved by a group of statistical methods experts, e.g. WGCATCH before implemented in the RDB.

ANNEX 4 WP2 Principles in the implementation of sampling designs

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

Deliverable 2.1 – Principles in the implementation of a sampling design

Lead: MSS, UK.

Core Team: Alastair Pout (MSS, UK)
Liz Clarke (MSS, UK)
Mary Christman (MCC, US)

Table of Contents

Principles in the implementation of a sampling design

1 Introduction

Outlined below are the guiding principles underlying a probability based regional sampling design for the collection of fisheries data, involving both on-shore or at-sea sampling programmes. These principles are based on the standard survey methodologies needed to make inferences about a population from a sample of the elements of that population. The implementation of such a design is set out as a series of steps, and it is assumed that in many regional designs that individual countries, or scientific institutions, are likely to be operating as strata within a regional design. It is appreciated that in fisheries data collection there are many challenges to the practical implementation of these statistical principles in the field, but as in any implementation every effort should be made to ensure that the sampling protocol is as close to an unbiased probabilistic design as possible.

Useful texts on sampling include Cochran (1977), Jessen (1978), Kish (1965), Lohr (2010), Thompson (2012), Särndal *et al* (1992). Moser and Kalton (1992) provide a non-technical appraisal of survey sampling in a social context, which in many ways is highly applicable to the situation found in fisheries sampling. Eurostat (2008) have produced Survey Sampling Reference Guidelines and the European Statistical System Quality Assurance Framework (2015) provides the context in which many European regional fisheries data collection designs are likely to be operating. Additionally the Global Adult Tobacco Survey (GATS) manuals provide a good example of the widespread implementation of a sampling design across diverse sampling strata, where the implementation is by autonomous institutions but the sampling practices are harmonised to common data collection and estimation principles. The work of ICES expert groups specifically related to sampling design can be found in ICES reports WKPRECISE (2009), WKMERGE (2010), WKPICS 1 (2011), WKPICS 2 (2012), WKPICS 3 (2013) and for at-sea sampling SGPIDS 1 (2011), SGPIDS 2 (2012), SGPIDS 3 (2013), and the ongoing work of WGCATCH (2014, 2015). Lumley (2010) demonstrates the use of the “survey” package in the R statistical language for estimation from complex survey designs.

2 Principal steps in the implementation of a sampling design

The Spatial, Temporal and Biological Scope of the Population

From the end users perspective, the population of interest is the set of individuals about which inference is desired. For example, it might be the catch of species of demersal fish in the North Sea. Hence, in this context, the population will be all of the individuals caught and either landed or discarded by fishing vessels. For this population, one or more characteristics such as the age, length, weight, maturity, are required to be estimated.

The *study* population are those elements of the population which are accessible or otherwise potentially available to be sampled. For example, while the population of interest may be all fish caught by national flag vessels, the study population may be subtly different being fish of the EU national flag fleets landed into ports in EU countries. The landings into non EU countries may be unavailable for sampling. In this case, the inferences can only be made about the sub-population of fish landed in the EU ports unless some assumptions are made about the fish landed in non-EU ports.

From the sampler's perspective, the statistical population is the set of elements on which measurements are made. For example, if information on the age structure of a stock is desired, then the element is the age of an individual fish. If total weight landed for a species is desired the element might be a box of sorted fish brought into port by a vessel.

These different facets of the population are important for identifying the accessible and available elements about which information is desired and obtainable.

Information Requirements

Determine exactly the purpose of the data collection. Is it to provide the age structure of the catch? Discard weight estimates? Species composition data? Estimates of the bycatch of marine mammals? The end users of the information that can be obtained from a sampling design are critical at this stage. Their input and requirements inform the overall design. For example, are estimates by gear type or country important (e.g. estimates of total catch associated with a particular métier)? Is the variability over the year important? In addition to the need for estimates of the desired population metrics, a statement about the achievable or desired precision of those estimates is also needed. The latter should be clearly related to a realistic assessment of the resources (person time) and finances available, and is likely to require discussion with the end-user. Schemes that are designed to address clearly defined requirements within clear budgets are far more likely to provide good estimates than schemes whose aims are poorly defined and inadequately funded.

The Sampling Frame

The sampling frame is the list of all elements that completely encompass the accessible component of the population. It is through this sampling frame that we gain access to the population and the list that is used when probabilistically selecting which elements in the population to sample. In most instances involving catch sampling over the course of a year (both at-sea and on-shore) a precise sampling frame does not exist a priori; the actual dates and locations of the landings by individual vessels, and the number and activity of vessels operating at-sea over the coming year is not known before they occur. For practical purposes however the sampling frames can be based on last year's logbooks and sales note data, and vessel registers, with an expectation that the temporal pattern of fishing activities for the coming year will be broadly similar to that of the preceding year. Depending on the sampling scheme different sampling frames will be used to encompass different fractions of the population; on-shore sampling frames are typically lists of ports, markets and processors where landings can be accessed, at-sea sampling frames are lists of vessels where the whole catch can be accessed. As much associated information as is possible can be obtained about the sampling elements in order to inform the stratification, likely clustering, multistage sampling, and similar needed to effectively obtain access to the ultimate sampling elements. This provides important information too about the suitability as sampling locations.

A sampling frame for small scale or recreational fisheries may not readily exist and will often have to be compiled from expert knowledge, or through the utilisation of publically available contact information available, e.g. general household surveys, angling club membership records, lists of holiday lodges, phone records, electoral registers etc.

The sampling frame was described above as that part of the population that is accessible and which can be listed for possible selection in a sample. Even when a sampling unit is identifiable it might not be "sample-able". For example, for on-shore sampling some ports might consign all landings directly to a market and so there is no opportunity for sampling; some processors may simply not allow access. For at-sea sampling some vessels will be unsuitable for carrying observers. These elements need to be retained in the sampling frame but will not be part of the sampling design and any data manipulations should include recognition of the effect of their exclusion from sampling.

Sampling Design

The most common and cost-effective sampling design for catch sampling is a multi-stage stratified random sampling design. Here the population is first divided into non-overlapping subgroups, called strata. Within each stratum, a multi-stage design is usually implemented as this approach recognizes and uses the nested structure of the elements that compose the population.

Stratification

Most fisheries sampling designs will be stratified. This involves splitting the entire sampling frame into a number of non-overlapping groups, the sampling elements of which constitute a unique subset of the sampling frame. Hence every sampling element should be included in a sampling stratum, and no sampling element should be listed in more than one stratum. It is important that the sampling stratum to which a sampling element belongs is clearly and unambiguously recognisable at the time of sampling.

The sampling strata can be used to group the more homogeneous elements together, as a means of increasing the precision of a final estimate (the "stratification principle" (Cochran, 1977)). Quite commonly however in fisheries sampling stratification is used as a means of meeting logistical demands of the implementation of the sampling design. These two purposes may or may not align. For example, if stratification is based on geographic regions, such as might occur if a single sampler is allocated to a set of spatially contiguous ports, this may meet the requirements for logistical efficiency but may or may not group ports with similar characteristics. Care should be taken to consider whether logistical stratification is actually necessary, and the effect it has on the estimates should be checked, for example by using simulation studies to compare different sampling designs.

The use of sampling strata enables sampling effort to be directed disproportionately *between* different strata, but each sampling unit *within* a stratum should, with a random selection design, have the same chance of being selected.

It should be noted that when stratifying in order to associate like elements with like within each stratum, there should be relatively few strata within a population. If too many strata are identified and the total available sample size is limited, then each stratum may have so few samples that the standard error of the stratum-level estimates are so large as to completely negate the advantage of stratifying elements into homogeneous groupings. When stratifying for logistical purposes, the number of strata is determined by the optimal allocation of resources in order to ensure good coverage of the population. The perceived benefits of this kind of stratification should be carefully assessed. Of course there is still the need to ensure that the number of strata is not so large as to reduce the sample sizes in each stratum to levels that would not allow precise estimation of the quantities of interest.

Multistage cluster sampling

For most if not all fisheries survey sampling multistage hierarchical cluster sampling will be involved in the selection of the actual fish or shellfish to measure. Cluster sampling is where the elements in the population are presented in distinct groups, such as vessels at a port, or fish within a basket, and a number of these can be sampled at the same time. In fisheries it is generally the case that the elements in these clusters are more similar to each other, than to sampling elements in the population as a whole. Multi-stage sampling is where there is a nesting of clusters in a hierarchical fashion, so that each level in the nested hierarchy has to be sampled in a sequential order, e.g. hauls within a trip, baskets from a haul, fish within a basket. The hierarchy and the clustering is often the result of the working practices of the industry and is thus determined by the exact situations encountered at the time of sampling.

Sampling hierarchy

The primary sampling units (PSU's) are first level of nesting at which sampling elements are available to be selected. In the case of fisheries sampling, the elements in the sampling frame are selected over a period of time hence the PSU commonly will be the combination of a sampling element, say port, and the time interval during when it is selected. It is important to have a clear understanding of what the PSU is for each sampling scheme. For on-shore sampling it will generally be a location (port market or processor) and a day. For at-sea sampling it will be vessel and fishing trip or a trip with a particular start-date, depending exactly on how sampling units are defined and selected.

Once the PSU is identified, the next step in the process is identifying sub-elements to be selected for sampling within the PSU. For example the multistage hierarchical cluster sampling for on-shore markets on a particular day (the PSU in this example) may involve selecting one or more vessels landing at the market. These would be the secondary sampling units (SSU's). Then, having selected a vessel, one or more boxes within the landings of a species (the tertiary sampling unit (TSU)), and finally sampling individual fish within a selected box. For at-sea sampling on a specific fishing trip, the hierarchical cluster sampling may involve sampling: hauls within a selected trip, baskets within selected hauls, finally individual fish within the selected baskets.

Sampling Effort

Having potentially identified the sampling design to be employed (identification of strata, identification of sampling units and the hierarchical structure of sampling units), the next step is to determine the amount of sampling that can be performed. This requires working out the allocation of visits to sampling locations by sampling teams over time (seasonally or annually) depending on the temporal nature of the fishing activity at the port or market (PSU). This requires that development of the sampling design and assessment of the amount of sampling to be performed is an iterative process. The sampling design is dependent on the resources available, for example in the first instance the person hours available for sampling, to ensure that what is planned can be achieved. A typical on-shore sampling design will specify the number of visits to specified sampling sites (ports, markets and/or processors) over a specified number of weeks. For at-sea sampling the design will specify the number of trips to be undertaken over a year.

Once the amount of sampling that can be done at the first stage is done, the number of secondary, tertiary, etc., units that can be sampled is based on similar considerations, for example the amount of person-hours available for sampling, the amount of time required to fully sample the selected units, and the time-window available for sampling.

Selection Protocols

At all the stages of a sampling hierarchy from the PSU down to the individual fish a formalised selection procedure needs to be set up. Thus whenever the sampling team has to make a choice of which elements to sample a selection protocol exists to determine and record that decision. The sampling team should not be selecting which samples to measure based on subjective judgements. These protocols can include printed selection sheets with randomised number strings or electronic recording devices with random selection routines. If verifiable methods can be devised, these are preferable and offer better guarantee of correct implementation. Selection protocols should be based on random or systematic random sampling as this ensures that the samples selected are unbiased and representative of all the possible samples that are available. Probability based selection also underpins the mathematical theory on which the estimates and measures of variance are calculated, hence without probability selection there is no reliable measure of the variability of the estimates derived from the samples. That said, random sampling or systematic random sampling is not always an easy thing to achieve in the field, and the exact protocols adopted will depend on the specifics of the sampling situation. The protocols have to be workable, and suitable recording sheets need to be researched and tested in the field prior to implementation.

Non-response and Refusals

With formalised sampling protocols and a documented selection process in place then non–response rates can be recorded. Non-response encompasses all the instances where elements that were selected for sampling did not provide data. The non-response rate provides an estimate of the proportion of the population from which data is actually obtained, and so potentially allows estimates to be adjusted to reflect the whole population. The extent to which the non-responsive elements in the population differ from those for which data are collected can also be explored to ascertain whether the data collected are biased. The refusal rate is a particular subset of non-response that applies to sampling units that have been contacted and refused to allow data to be collected. In the fisheries context with at-sea sampling the refusal rate provides a useful measure of the level of access observers have to the vessels of a fleet. In the on-shore context it can for example reflect the accessibility of fish on a market. Collecting non-response data is both informative as to the quality of the estimates and facilitates the improvements that can be made to the operation of sampling schemes.

Inclusion Probabilities and Sample Weights

Each of the stages in this multistage hierarchy will generate an inclusion probability, the probability of being included in the sample, which at its simplest is the ratio of the *n* elements selected from the total N available. The inverse of the combined inclusion probabilities over all stages in the sampling hierarchy generates the sample weight. The sample weight can be usefully considered to represent the total number of individuals in the particular population being sampled that the measured sample represents and, in order to derive unbiased estimates, the sampling weight must be correctly calculated. The correct sample weight is thus an integral part of the collected sample and reflects the circumstances under which the sample was collected.

Estimation

The generation of estimates of population parameters is the whole reason for survey sampling. Hence the estimation process to be used should be determined in association with the design of the survey so that the sampling design, selection protocols, and data to be collected can be set up accordingly. The most widely used and versatile design-based estimator for stratified probability based sampling is the Horvitz-Thompson estimator (Horvitz & Thompson, 1952). This estimator is applicable to many, if not all, of the design stages and protocols outlined here that are based on probability based selection of sampling units. Lumley (2010) outlines the use of the Horvitz-Thompson estimator in the “survey” package of the R statistical language.

Ratio Estimators

In situations where N is not known it can be possible to estimate the sampling probability using a ratio estimator based on an auxiliary variable that is related to the “size” of the population being sampled; historically weight has been commonly used in fisheries sampling. However, the choice of auxiliary variables is not always straightforward as their relationship with the variable being measured needs to be well understood. Inappropriate auxiliary variables can lead to biased estimates.

Data Recording and Storage

Much fisheries sampling data ends up being archived in electronic databases of one form or another, and the critical elements of this process are to ensure that all pertinent data can be stored in a readily assessable fashion, and that the data model underpinning the data base does not impose inappropriate restrictions on what data can be recorded, how it is stored, and how it is accessed. The basic principle of storing raw data as it was collected (and not derived data that has been calculated) should be adhered to. Another basic aim of all scientific sampling is to faithfully record the variability that exists in the sampled population so, while there are legitimate outlier and error trapping routines, these should not be so draconian as to violate that principle.

Optimization of Sampling Designs

Optimization of a sampling design has several aspects, including (1) maximizing the efficiency of the samplers available for taking the measurements, (2) maximizing the efficiency of the statistical estimates of the desired parameters, and (3) finally maximizing the usefulness of the information obtained from the sampling. The sampling designs for fisheries data collection have to be envisaged to run over the medium to long term in order to provide consistency of data collection and of the derived estimates. It needs to be remembered that it is difficult to optimise sampling schemes that seek to provide data to meet a wide range of potentially different data needs.

For (1), this may require structuring sampling to minimize travel among PSUs and setting temporal schedules for visiting PSUs such that samplers are working on a regular schedule. This is often accomplished by structuring geographical strata so that samplers are visiting ports on a regular schedule during the fishing seasons. Visits to sampling locations can be organised in a random or systematic random way across the time period, and will allow practical and sensible decisions to be made about the best use of sampling teams and the coverage of the fishing operations. This requires that the strata boundaries be set based on some measure of the “size” or diversity of the sampling strata. This may be the number of trips, the landed weight the number of species etc. The cost of the sampling may also be a consideration.

For (2), the variance of the estimates is related to choice of stratum boundaries and inversely related to the number of observations taken within each stratum. Adhering to the stratification principle will ensure lower variance of the estimates but in most instances this is not completely possible due to the requirements of the availability of samplers and the wide range of likely data needs. Hence, the number of samples taken within each stratum becomes more important as it is the main method by which variance can be reduced. The best way to accomplish lower variance is to allocate sampling effort in such a way that more effort is expended in the strata that are more variable (for example, ports/markets that are highly variable is the types and sizes of fish being landed). The best approach here is to review the results of prior years of sampling to identify those sampling design aspects and sample sizes that should be modified to increase efficiency.

Item (3) is more difficult to implement without expert knowledge of the study population(s) and end user input into the sampling design. Here the statistician designing the study must ensure that the data requirements of the end users is met or can be met with additional input from other areas. For example it may involve the collection of additional auxiliary information from the sampling elements that allows estimates to be generated for “domains of interest” from particular subsets of the realised sampling data.

References

Cochran, W.G. (1977) Sampling Techniques. John Wiley and Sons, New York.

Eurostat 2008. Survey Sampling Reference Guidelines. http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-08-003

Eurostat 2015. Quality Assurance Framework of the European Statistical System. http://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646

Global Adult Tobacco Survey Collaborative Group. Global Adult Tobacco Survey (GATS): Sample Design Manual Version 2.0 Atlanta, GA: Centres for Disease Control and Prevention, 2010.

Global Adult Tobacco Survey Collaborative Group. Global Adult Tobacco Survey (GATS): Sample Weights Manual Version 2.0 Atlanta, GA: Centres for Disease Control and Prevention, 2010.

Horvitz, D. G.; Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *Journal of the American Statistical Association*, 47, 663–685.

ICES 2009. Report of the ICES Workshop on methods for to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE). ICES CM 2009 / ACOM:40

ICES 2010. Report of the ICES Workshop on methods for merging métiers for fishery-based sampling (WKMERGE). ICES CM 2010 / ACOM:40

ICES. 2011. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS 1). ICES 2011 /ACOM:50

ICES 2012. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS 2). ICES 2012 /ACOM:50

ICES 2013. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS 2). ICES 2012 /ACOM:56

ICES 2011. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS 1), 8-10 November 2011, Bilbao, Spain. ICES CM 2011/ACOM:52

ICES 2012. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS 2), 6-9 November 2012, Copenhagen. ICES CM 2012/ACOM:54

ICES 2013. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS 2), 6-9 November 2012, Copenhagen. ICES CM 2012/ACOM:54

ICES 2014. Report of the Workshop on Commercial Catches (WKCATCH2), 10-14 November 2014, Copenhagen. ICES CM 20142/ACOM:36

Jessen, R. A. 1978. Statistical survey techniques, John Wiley and Sons, New York.

Kish, L. 1965. Survey Sampling. John Wiley & Sons, New York.

Lohr, S. 2010. Sampling: design and analysis 2nd Ed. Brooks/Cole, Boston.

Lumley, T. 2010. Complex surveys, a guide to analysis using R. John Wiley & sons New Jersey.

Moser, C.A. Kalton, G. 1979. Survey Methods in Social Investigation. Gower Publishing, Aldershot.

Särndal, C.-E., Swensson, B., Wretman, J. 1992. Model Assisted Survey Sampling. Springer-Verlag, New York.

Thompson, S.K. 2012. Sampling. 3rd Ed. John Wiley & Sons, New York.

ANNEX 5 WP2 Data sharing agreement

fishPi project (MARE/2014/19)

“STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION”

WP2 – Regional sampling design for commercial fisheries

WP2 – Data Sharing Agreement

Lead: MASTS, Scotland

Core Team: Mark James, Emma Defew (MASTS, Scotland)

Data sharing agreement

Here we reproduce the data sharing agreement which was drawn up by MASTS and circulated to all project partners to be signed by the heads of the respective scientific institutions and administrations prior to the request for logbook and sales note data. This data sharing agreement was effective from 10th April 2015. The agreement specified the usage of the data and the protocols for storage; the latter including a password protected limited access area on the project SharePoint. The SharePoint was provided by ICES

DATA SHARING AGREEMENT

This Data Sharing Agreement (hereinafter referred to as “the Agreement”) is dated 10th April 2015, hereinafter referred to as “the Effective Date”

AMONG:

1. **THE UNIVERSITY COURT OF THE UNIVERSITY OF ST ANDREWS**, a charitable body registered in Scotland under the registration number SC013532 and incorporated by the Universities (Scotland) Act 1889, as amended by the Universities (Scotland) Act 1966, and having its principal office at College Gate, North Street, St Andrews, Fife KY16 9AJ the ‘Lead Partner’

2. **THE SCOTTISH MINISTERS ACTING THROUGH MARINE SCOTLAND**, Science, Marine Laboratory, 375 Victoria Road, Aberdeen, Scotland, AB11 9DB

3. **THE SECRETARY OF STATE FOR ENVIRONMENT FOOD AND RURAL AFFAIRS** acting through the Centre for Environment, Fisheries and Aquaculture Science (Cefas) Headquarters (Cefas) – Pakefield Road, Lowestoft Suffolk NR33 0HT, UK. Cefas is an Executive Agency of the UK Government Department – Defra

4. **INSTITUTE FRANCAIS DE RESEARCHE POUR L’EXPLOITATION DE LA MER**, a public institute of industrial and commercial nature (VAT number FR46 330 715 368) Avenue du Général de Gaulle, BP32, Port en Bessin, France, 14520

5. **JOHANN HEINRICH VON THÜNEN-INSTITUTE, FEDERAL RESEARCH INSTITUTE OF RURAL AREAS, FORESTRY AND FISHERIES**, Bundesallee 50, 38116 Braunschweig, Germany Here acting: Thünen- Institute of Sea-Fisheries, Palmaille 9, 22767 Hamburg, Germany

6. **TECHNICAL UNIVERSITY OF DENMARK** Anker Engelunds Vej 1, Bygning 101A, 2800 Kgs. Lyngby, Denmark.

7. **SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES**, Turistgatan 5, Lysekil, Sweden, SE-453 30

8. **STICHTING DIENST LANDBOUWKUNDIG ONDERZOEK**, a foundation registered in The Netherlands under the registration number 09098104, and having its principal office at Droevendaalsesteeg 4, WAGENINGEN 6708 PB, The Netherlands.

9. **INSTITUTE FOR AGRICULTURAL AND FISHERIES RESEARCH**, a Flemish Scientific Institute to the Policy Domain “Agriculture And Fisheries” of the Flemish Government, Burg. van Gansberghelaan 92, Merelbeke, Belgium, 9820

10. **INSTITUTO PORTUGUÊS DO MAR E DA ATMOSFERA, INSTITUTO PUBLICO** (IPMA, I.P.), sede na Rua C ao Aeroporto, 1749-077 Lisboa, Portugal com o registo PT510 265 600.

11. **INSTITUTO ESPAÑOL DE OCEANOGRAFÍA** (IEO), a public research body, located at C/ Corazón De María 8, 28002, Madrid, Spain, with VAT number ESQ2823001I

12. **FUNDACION AZTI**, a non-profit private foundation, Txatxarramendi ugartea, Z/G, Sukarrieta, Bizkaia, Spain, 48395

13. **INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA**, an intergovernmental marine science organization founded in 1902, hosted by the country of Denmark and registered under the registration number 12063814, and having its principal office at H.C. Andersens Boulevard 44-46, 1553 Copenhagen, Denmark

hereinafter, jointly or individually, referred to as **"Parties"** or "**Party**"

Relating to the Action entitled **"Strengthening regional cooperation in the area of fisheries data collection in the North Sea and Eastern Arctic"** in short MARE/2014/19 and hereinafter referred to as "**Project**"

WHEREAS:

(A) The Parties have all entered into the Consortium Agreement (as defined below) in relation to the Project.

(B) In order to carry out the Project, the Parties require to share the Data (as defined below).

(C) The Data (as defined below) is of a disaggregated and confidential nature and so the Parties wish to specify or supplement binding commitments among themselves in addition to the provisions of the specific Grant Agreement (as defined in the Consortium Agreement) and the Consortium Agreement (as defined below).

NOW, THEREFORE, IT IS HEREBY AGREED AS FOLLOWS:

1 Definitions

1.1 The following definitions will apply to this Agreement: -

1.1.1 **"Consortium Agreement"** means the consortium agreement entered into by the Parties dated 10 April 2015 in relation to the Project ;

1.1.2 **"Data"** means all data shared by the Parties under this Agreement as specified in Part 1 of the Schedule;

1.1.3 **"Depersonalised Data"** means information that relates to individuals where it is not possible to identify individuals from that information, whether in isolation or in conjunction with any other information;

1.1.4 **"Non-Personal Data"** means information that does not relate to people including information about organisations, resources, projects or information about people that has been aggregated to a level that is not about individuals;

1.1.5 **"Security Policy"** means the security policy as set out in Part 2 of the Schedule.

1.2 Any words or phrases with capitalised letters shall have the meaning ascribed to them in the Consortium Agreement.

1.3 This Agreement is supplementary to the Consortium Agreement. In the event of a conflict in the application or interpretation of terms set out in this Agreement and the Consortium Agreement, the terms of the Consortium Agreement shall take precedence.

2 Data Sharing

2.1 The Parties agree to share the Data as described in Part 1 of the Schedule. .

2.2 The Parties undertake that they will only transfer Depersonalised Data and Non-Personal Data under this Agreement.

2.3 The Parties undertake that they will only use the Data in relation to the Project and for the purposes specified in Part 2 of the Schedule.

3 Security

3.1 The Parties agree to comply with the Security Policy and User Access and Use Policy as set out in Part 2 of the Schedule later in this document.

4 Duration

4.1 This Agreement shall commence on the Effective Date and shall continue for the duration of the Project. The existence of this Agreement is subject always to the Consortium Agreement remaining in full force and effect. In the event that the Consortium Agreement terminates, this Agreement will automatically terminate.

5 Confidentiality

5.1 The Parties agree and acknowledge that the Data shared under this Agreement will be regarded as Confidential Information under the Consortium Agreement and the Confidentiality provisions in section 10 of the Consortium Agreement will apply.

6 Liability

6.1 Any breach of this Agreement by a Party shall be deemed to be a breach of the Consortium Agreement and the parties liability to each other will be as set out in Section 5 of the Consortium Agreement and the Parties shall have all rights and remedies available to them under the Consortium Agreement in relation to the same.

7 Withdrawal of a Party

7.1 If any Party wishes to withdraw from this Agreement, they must follow the procedure set out in the Consortium Agreement.

8 Assumption of a new Party

8.1 If any new party requires to become party to this Agreement, the Lead Partner will, on behalf of the Parties to this Agreement, if directed to do so by the Project Management Committee, assume such new party as a Party to this Agreement by entering into an accession agreement which requires them to comply with the terms of this Agreement.

9 Execution in Counterpart

9.1 This Agreement may be signed in counterpart.

10 Miscellaneous

10.1 If any party to this agreement is authorised under the Consortium Agreement to grant a third party access to the Data for the purpose of performing their obligations under the Consortium Agreement, that party will ensure that any such third party complies with the terms of this Agreement. No other third parties should be granted access to the Data without the Project Management Committee's prior written agreement and subject always to them signing up to the terms of this Agreement in accordance with the terms of clause 8.1.

10.2 Should any provision of this Agreement become invalid, illegal or unenforceable, it shall not affect the validity of the remaining provisions of this Agreement. In such a case, the Parties concerned shall be entitled to request that a valid and practicable provision be negotiated which fulfils the purpose of the original provision.

10.3 Except as otherwise provided in Section **Error! Reference source not found.** of this Agreement, no Party shall be entitled to act or to make legally binding declarations on behalf of any other Party or of the consortium. Nothing in this Agreement shall be deemed to constitute a joint venture, agency, partnership, interest grouping or any other kind of formal business grouping or entity between the Parties.

10.4 Any notice to be given under this Agreement shall be in writing to the addresses and recipients as listed in the most current address list kept by the Lead Partner.

10.5 If it is required in this Agreement that a formal notice, consent or approval shall be given, such notice shall be signed by an authorised representative of a Party and shall either be served personally or sent by mail with recorded delivery or telefax with receipt acknowledgement.

10.6 Other communication between the Parties may also be effected by other means such as e-mail with acknowledgement of receipt, which fulfils the conditions of written form.

10.7 Any change of persons or contact details shall be notified immediately by the respective Party to the lead partner. The address list shall be accessible to all concerned.

10.8 No rights or obligations of the Parties arising from this Agreement may be assigned or transferred, in whole or in part, to any third party without the other Parties' prior formal approval.

10.9 Amendments and modifications to the text of this Agreement, not explicitly listed in this Agreement, require a separate written agreement to be signed between all Parties.

10.10 Nothing in this Agreement shall be deemed to require a Party to breach any mandatory statutory law under which the Party is operating.

10.11 This Agreement is drawn up in English, which language shall govern all documents, notices, meetings, arbitral proceedings and processes relative thereto.

10.12 This Agreement shall be construed in accordance with and governed by the laws of Belgium excluding its conflict of law provisions.

10.13 The parties shall endeavour to settle their disputes amicably.

10.14 Any dispute, controversy or claim arising under, out of or relating to this contract and any subsequent amendments of this contract, including, without limitation, its formation, validity, binding effect, interpretation, performance, breach or termination, as well as non-contractual claims, shall be submitted to mediation in accordance with the WIPO Mediation Rules. The place of mediation shall be Brussels unless otherwise agreed upon. The language to be used in the mediation shall be English unless otherwise agreed upon. If, and to the extent that, any such dispute, controversy or claim has not been settled pursuant to the mediation within 60 calendar days of the commencement of the mediation, the courts of Brussels shall have exclusive jurisdiction.

AS WITNESS: The Parties have caused this Agreement to be duly signed by the undersigned authorised representatives in separate signature pages the day and year first above written.

Here followed the signatures of the parties

This is the Schedule referred to in the foregoing Data Sharing Agreement entered into by the University Court of the University of St Andrews in relation to the Project entitled "Strengthening regional cooperation in the area of fisheries data collection in the North Sea and Eastern Arctic"

PART 1

THE DATA

The data required to design the sampling scheme under the Consortium Agreement are highly disaggregated and confidential.

In preparing the project proposal, all project participants were made aware in writing and have in principle agreed to provide access (where possible) to trip-level log-book and sales note data with the caveat that these data will be anonymised, shared and used among project participants for the purpose of sampling design within the time scale of the project.

For the avoidance of doubt, these Data will not be available for dissemination after completion of the project.

These Data are used by most institutes to compile annual data submissions for stock assessment and fisheries management during April and May of each year, and should therefore be available for processing in the required format by the time they are needed for the start of the project.

Data provided will be in the form of one or more of the following:

- Anonymised trip-level landings information
- Anonymised trip-level sales information
- Anonymised catch sampling data and PETS sampling data
- RV data on stomach sampling
- Small scale fisheries
    - fleet summaries (no of vessels in fleet register)
    - anonymised landings information (e.g monthly declarative forms) where available
    - anonymised sales information
- Recreational fisheries
    - meta-data on recreational fisheries

The security, access and use of any Data supplied will be as set out in this agreement and more specifically described in Part 2 of the Schedule to this Agreement.

PART 2 - THE SECURITY POLICY AND ACCESS TO AND USE OF DATA

The Parties agree and undertake that they will comply with the terms of this Security Policy and the terms setting out use of and access to the Data.

Use of the Data

The Data can only be used for the purposes of the Project

The Data can only be used during the duration of the Project.

Any publication of data or analysis derived from Data must be authorised by the Project Management Committee (as defined in the Consortium Agreement) and under such authorship as determined by the Project Management Committee.

The Parties agree that all publications will be anonymous in order to protect the source of the Data.

The Data should be used for research and analysis purposes accordance with the Consortium Agreement and for no other purposes. The raw Data in the form provided by the Parties as per Part 1 of the Schedule should not be published in its raw state.

All Data contributed by another party will be deleted from Sharepoint and shall not be accessed or used by a party following expiry or termination of the Consortium Agreement.

How the Data will be held/accessed

The parties shall upload the Data onto Sharepoint under the direction of the Project Management Committee. Each party will be provided with secure passwords and credentials to allow them to access the Data. The Parties each undertake that they will keep their passwords and credentials secure and that only project members **approved by the Project Management Committee** that require to access the Data in order to carry out the Project will be granted access to the data. Downloads should be kept securely and in accordance with clause 10 of the Consortium Agreement.

Data shall be held securely in accordance with the ICES Data Policy http://www.ices.dk/marine-data/guidelines-and-policy/Pages/ICES-data-policy.aspx and INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA agrees that it shall not make the relevant data publically available.

Security Breaches

1 If any Party:

1.1 becomes aware of any unauthorised or unlawful processing of any Data or that any Data is lost or destroyed or has become damaged, corrupted or unusable;

1.2 becomes aware of any breach of this Security Policy; or

1.3 learns or suspects that any unauthorised third parties have obtained any Data,

that Party shall, at its own expense, promptly notify the Project Management Committee and fully co-operate with it to remedy the issue as soon as reasonably practicable.

2 The Parties agree to co-operate with the Project Management Committee's reasonable security investigations.

ANNEX 6 WP2 Case study data format

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2 – Regional sampling design for commercial fisheries

Deliverable 2.1 – Case study data format

Lead: MSS, Scotland

Core Team: Alastair Pout, Liz Clarke (MSS, Scotland)
José Castro (IEO)
Marie Storr-Paulsen (DTU Aqua, Denmark)
Edwin van Helmond (IMARES, Netherlands)

fishPi WP2.3 Case Study Data Request

As part of the fishPi project WP2, we request the following trip-level landings and sales information from your institute, for the regional sampling design case studies in which you are taking part.

We would like you to prepare this data ideally by 22nd June 2015 in time for the WP 2 start-up meeting. It can be analysed within national labs but it cannot be shared until your institute has signed the data sharing agreement DSA. Please make every effort to see that the DSA is signed on time. Data are to be uploaded to the fishPi SharePoint https://community.ices.dk/fishPi/Data/ by the individual designated by the national institution, and accepted by the fishPi PMC.

Data are requested for the years 2013 and 2014.

The scope of the data call is set out in part 1 below and defined in terms of the vessels, the areas and the target species we would like data for. Please provide all possible data that meets the scope of these requests for both on-shore and at-sea sampling, as partial or incomplete data sets will invalidate any possible regional design.

Please also note that we require data to be provided by individual fishing trip for all trips that meet the scope of the call for each case study. For these trips we need the *full* trip information which includes *all* other landed species from those trips. The information required and the format to be used is set out in part 2 below.

These data are to be used to design on-shore and at-sea regional sampling schemes. This will involve the construction of suitable sampling frames of on-shore locations for on-shore sampling, and vessel lists for at-sea sampling. It is envisaged that sales note data and logbook data will be suitable for this purpose.

For on-shore sampling designs the sale location provided with sales note data can often be the best means of identifying a suitable on-shore sampling location, If landing location is from provided from logbook data then please try to ensure that these locations are suitable for on-shore sampling.

For at-sea sampling designs data from the trips of individual vessel's logbook records will be well suited for the task. Please ensure that all relevant vessels in national fleet's area included, that all the trips can be identified to one of these vessels, and that individual vessels have the same vessel id for the two years of data.

All data supplied will be used and stored according to the terms of the data sharing agreement, in particular all vessels are to be anonymised, and all lengths are to be converted to length classes.

Many Thanks

Marie Storr-Paulsen, Alastair Pout, Liz Clarke, Edwin van Helmond, José Castro Pampillón

Part 1 Scope

Case Study 1 Small Pelagic fisheries
Vessels: Vessels using pelagic fishing methods. (i.e. single and twin midwater trawls and pelagic seines; gears OTM, PTM, PS)
Areas: All ICES area in FAO area 27
Target species: herring, mackerel and sprat

Case Study 2 North Sea Demersal
Vessels: Vessels using demersal fishing techniques (gillnets, hooks & lines, demersal trawls and demersal seines, beam trawls, i.e. all gears defined in vessel type (see below) with the exception of DRB FPOand TM regardless of the métier defined target assemblage
Areas: ICES divisions IIIa, IV, VIa, VIb & VIId
Target species: all species of fish and *Nephrops*, but no other shellfish

Case Study 3 Southern North Sea flatfish
Vessels: Vessels using demersal fishing techniques (gillnets, hooks & lines, demersal trawls and demersal seines, beam trawls)
Areas: ICES divisions IIIa, IV, & VIId
Target species: plaice, sole, turbot & brill
Note that the data for case study 3 is a subset of the data for case study 2, hence there is no need to provide CS3 data IF you are providing data to CS2.

Case Study 4 Hake
Vessels: Vessels using demersal fishing techniques (mainly trawls, set gillnets and set longlines)
Areas: Division IIIa, ,Subareas IV, VI, VII and VIII, and Division IXa
Target species: hake

Part 2

Data should be provided in the following format as a R data frame or a comma separated CSV flat file consisting of 25 columns. The first 12 columns should be identical and repeated for any particular trip, from column 13 onwards a new row should be started for each fish species recorded on the trip.

Field	Name	CodeLists	Comments
1	recType	One of: "on-shore scheme", "at-sea scheme" or "both"	"at-sea scheme" should be added if the record is only suited to an at-sea scheme and has no suitable "no-shore sampling location"for column 17. "on-shore scheme" should be added if the record is only suited to an on-shore scheme and has no "at-sea sampling location" for column 16. "both" should be recorded if the record can be used for both on-shore and at-sea schemes and both sampling locations, and columns 16 and 17 are both populated.
2	vslFlgCtry	ISO 3166 – 1 alpha-3 code	Vessel flag country
3	voyageId	Institute-specific code	Unique trip code within a country (need not be numeric)
4	vslId	Institute-specific code	Unique vessel id within a country (This need not be numeric but does need to be the same for the same vessel over time (i.e. not change between years).)
5	vslLenCls	DCF LOA classes VL0010 VL1012	DCF vessel overall length classes 0-10,10-12,12-18,18-24,24-40,40+ Please use the fleet segment code list (http://datacollection.jrc.ec.europa.eu/wordef/fleet-segment-dcf)
6	vslPower	<221KW,>221KW	For beam trawl fleets.
7	vslType		Predominant fishing gear as used in the DCF fleet segmentation: http://datacollection.jrc.ec.europa.eu/wordef/fleet-segment-dcf This should be unique for each vessel. The dominance criteria shall be used to allocate each vessel to a segment based on the number of fishing days used with each gear. If a fishing gear is used by more than the sum of all the others (i.e. a vessel spends more than 50 % of its fishing time using that gear), the vessel shall be allocated to that segment. If not, the vessel shall be allocated to the following fleet segment: (a) 'vessels using Polyvalent active gears' if it only uses active gears; (b) 'vessels using Polyvalent passive gears' if it only uses passive gears; (c) 'vessels using active and passive gears'.

DFN	=	Drift and/or fixed netters
DRB	=	Dredgers
DTS	=	Demersal trawlers and/or demersal seiners (includes OTB, OTT, PTB and SSC)
FPO	=	Vessels using pots and/or traps
HOK	=	Vessels using hooks
MGO	=	Vessel using other active gears
MGP	=	Vessels using polyvalent active gears only
PG		Vessels using passive gears only for vessels < 12m
PGO	=	Vessels using other passive gears
PGP	=	Vessels using polyvalent passive gears only
PMP	=	Vessels using active and passive gears
PS	=	Purse seiners
TM	=	Pelagic trawlers
TBB	=	Beam trawlers

8	depDate	YYYY-MM-DD	Departure date for the trip.
9	arvDate	YYYY-MM-DD	Arrival date for the trip.
10	Homeport	UN/LOCODE	Port in which the vessel is registered.
11	depLoc	UN/LOCODE	Departure port, - include ports in all countries
12	arvLoc	UN/LOCODE	Arrival port – include ports in all countries
13	sppCode	WoRMS	For all landed spp that are recorded on the trip (either from logbook or sales notes). Not recorded

		code	discarded species. WoRMS AlphiaID 6 digit species code http://www.marinespecies.org/index.php e.g. hake = 126484
14	sppName	WoRMS Scientific name	Accepted scientific name according to the WoRMS list http://www.marinespecies.org/index.php e.g. hake = *Merluccius merluccius*
15	landWt	Kg	Landed weight in kg for each species recorded in the landing/sales record (this will differ by species)
16	atSeaSampLoc	UN/LOCODE	The location in which an at-sea observer would be able to access the trip for at-sea sampling.
17	onShoreSampLoc	UN/LOCODE	The location in which an on-shore observer would be able to access the landed fraction of the catch from the trip (this may differ by species).
18	Rect	RDB list	ICES rectangle for each species recorded in the landing/sales record (this may differ by species).
19	Area	FAO area codes	ICES subarea, division or subdivision Please provide at the finest resolution possible using the FAO area codes. e.g. ICES subdivision Iva is 27.4.a (this may differ by species).
20	foCatEu6	RDB 2015 list	EU level 6 métier (this may differ by species).
21	fleetNat	Institute-specific code	National fleet definition (The national classification of vessels which is relevant to the sampling)
22	onShoreStratNat	Institute-specific code	Current National design stratification for on-shore sampling i.e for this record: which sampling strata would be currently be used
23	atSeaStratNat	Institute-specific codes	Current National design stratification for off-shore sampling i.e for this record: which sampling strata would be currently be used.
24	domainNat1	Institute-specific codes	Any current National estimation domain e.g. a domain for which this record may contribute to the estimate
25	domainNat2	Institute-specific codes	Any current National estimation domain e.g. another domain for which this same record may contribute to the estimate

Notes

Species - All species landed from the trip should be recorded hence each trip record will have the same number of rows as the number of landed species; the first 12 columns of the data will be duplicate records, the following columns will recorded species specific details such as rectangle and métier, that may differ between species.

Species - codes these should be in the 6 digit WoRMS codes

Species names – these should be the corresponding Scientific name as accepted in the WoRMS lists.
domainNat1 domainNat2 - these fields should be used to record the domains for which this record would be used at the national level. This is up to the data providing nation but the domains may be a single species stock assessment, an STECF effort category, an inshore fishery impact assessment etc; i.e. any relevant national usage of this data. Providing such information will enable the case study to evaluate how effectively a regional scheme could meet national data provision requirements.

ANNEX 7 WP2 Case study data compilation code

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2 – Regional sampling design for commercial fisheries

Deliverable 2.1 Case study data compilation code

Lead: MSS, Scotland

Core Team: Alastair Pout, Liz Clarke (MSS, Scotland)
Mary Christman (MCC, USA)
Ana Ribeiro Santos, Jon Elson (CEFAS, UK England)
Patrik Börjesson (SLU Sweden)
Kirsten Birch Harkannson, Marie Storr-Paulsen (DTU Aqua, Denmark)
Laurent Dubroca, (IFREMER, France)
José Castro, José Rodríguez (IEO, Spain)
Lucia Zarauz, (AZTI, Spain)
Chun Chen (IMARES, Netherlands)
Nuno Prista (IPMA, Portugal)
Julia Wischnewski (TI-SF Germany)

Data compilation and checking code

This document is intended to describe the process for the data cleaning of the files submitted by each Member State and the process undertaken to standardize the data submitted across all countries.

This is an iterative report created, using R Markdown and 'Knitr'. R Markdown is an authoring format that enables easy creation of dynamic documents from R. It combines the core syntax of markdown (an easy to write plain text format) with embedded R code chunks that are run so their output can be included in the final document. R Markdown documents are fully reproducible (they can be automatically regenerated whenever underlying R code or data changes).

The report highlights the issues and potential errors with the original dataset that may need to be resolved. If no issues or errors are found after running this code, the cleaned and standard dataset is produced that can be combined with the data from other countries and used for each case study.

The following report is an example of a data compilation report, using data submitted by England for Case Study 2 - North Sea Demersal Fisheries.

The code starts with setting up the working directory, import the necessary libraries, reference tables and the original dataset submitted by each country.

```
# Whatever directory you wish to put the project in
#myDir <- '~/2. R/R-project/fishPi/WP2/'
#Clear any previous calculation which may interfere in the current one
rm(list=ls())

#Set up the working directory
MyDrive<-"C:"
wd<-file.path(MyDrive,'Users', 'AS15', 'Documents', 'C6604', 'Data_Storage',
              'ARS', 'CS_quality_checks'); setwd(wd) # wd

data.dir<-file.path(getwd(),'data')#data
lookup.dir<-file.path(getwd(),'lookup')#figures
code.dir <- file.path(getwd(),'code') # code
output.dir <- file.path(getwd(),'output')

##### Add the relevant libraries
library(plyr)
library(dplyr)
library(reshape2)
#library(fishPiCodes)

##### lookup tables
load(file.path(lookup.dir,'RefTblDCFGears.RData')) # lookup table for vesseltypes
df2 <- readRDS(file.path(lookup.dir,'metiersAreas.rds')) # lookup table for metiers in each area
load(file.path(lookup.dir,'UNLOCODE.rda'))  # lookup table for location codes and positions
```

```r
##### input data
cs1 <- read.csv (file.path(data.dir, 'ENG_CS2.csv'))

#List the object in the environment
ls()
```

```
## [1] "code.dir"      "cs1"          "data.dir"      "df2"
## [5] "lookup.dir"    "MyDrive"      "output.dir"    "RefTblDCFGears"
## [9] "UNLOCODE"      "wd"
```

```r
#vslLenCls Function
#The vessel length function was created to normalize vessel length format
#across countries.

f.vslLenCls <- function(x)
{
  standard <- c("VL0010","VL1012","VL1218","VL1824","VL2440","VL40XX");
  if (!all(x$vslLenCls %in% standard))
  {
    new.vslLenCls <- as.vector(x$vslLenCls);

    n1 <- which(grepl("-<", new.vslLenCls));
  new.vslLenCls[n1] <- paste("VL", gsub("-<", "", new.vslLenCls[n1]), sep="");
    n2 <- which(grepl("<", new.vslLenCls));
    new.vslLenCls[n2] <- paste("VL00", gsub("<", "", new.vslLenCls[n2]), sep="");
    n3 <- which(grepl(">", new.vslLenCls));
    new.vslLenCls[n3] <- paste("VL", gsub(">", "", new.vslLenCls[n3]), "XX", sep="");

    x$vslLenCls <- new.vslLenCls;
  }
  return(x)
}
```

```r
#Check and standardise headings
names(df1)[names(df1)=="voyageld"] <- "voyageId"
names(df1)[names(df1)=="vslld"] <- "vslId"

df1 <- df1[,c ( "recType", "vslFlgCtry", "voyageId", "vslId", "vslLenCls", "vslPower",
                "vslType", "depDate","arvDate", "homePort","depLoc","arvLoc","sppCode",
                "sppName", "landWt", "atSeaSampLoc","onShoreSampLoc",  "rect" ,"area",
                "foCatEu6", "fleetNat", "onShoreStratNat", "atSeaStratNat","domainNat1",
                "domainNat2")]
```

Field headings standardised!

Check the Flag countries in the dataset

```
print (unique (df1$vslFlgCtry))
```

```
## [1] GBN GBE GBR GBS GBW GBI
## Levels: GBE GBI GBN GBR GBS GBW
```

Convert all the variables to character and numberic

```
#Put all the variables as character or numeric
df1$recType <- as.character(df1$recType)
df1$vslFlgCtry <- as.character(df1$vslFlgCtry)
df1$voyageId <- as.character(df1$voyageId)
df1$vslId<- as.character(df1$vslId)
df1$vslLenCls<- as.character(df1$vslLenCls)
df1$vslPower<- as.character(df1$vslPower)
df1$vslType<- as.character(df1$vslType)
df1$depDate<- as.character(df1$depDate)
df1$arvDate<- as.character(df1$arvDate)
df1$homePort <- as.character(df1$homePort )
df1$depLoc<- as.character(df1$depLoc)
df1$arvLoc<- as.character(df1$arvLoc)
df1$sppCode<- as.character(df1$sppCode)
df1$sppName<- as.character(df1$sppName)
df1$landWt<- as.numeric(df1$landWt)
df1$atSeaSampLoc <- as.character(df1$atSeaSampLoc )
df1$onShoreSampLoc<- as.character(df1$onShoreSampLoc)
df1$rect<- as.character(df1$rect)
df1$area <- as.character(df1$area)
df1$foCatEu6<- as.character(df1$foCatEu6)
df1$fleetNat<- as.character(df1$fleetNat)
df1$onShoreStratNat<- as.character(df1$onShoreStratNat)
df1$atSeaStratNat<- as.character(df1$atSeaStratNat)
df1$domainNat1<- as.character(df1$domainNat1)
df1$domainNat2<- as.character(df1$domainNat2)
```

All variables now character and numeric!

Standardize date format and check the date range

```r
df1$newdate <- strptime(as.character(df1$depDate), "%d/%m/%Y")
df1$txtdate <- format(df1$newdate, "%Y-%m-%d")
df1$depDate <- ifelse(is.na (df1$txtdate), df1$depDate, df1$txtdate)
df1$depDate <- as.Date(df1$depDate,format='%Y-%m-%d')

df1$newdate <- strptime(as.character(df1$arvDate), "%d/%m/%Y")
df1$txtdate <- format(df1$newdate, "%Y-%m-%d")
df1$arvDate <- ifelse(is.na (df1$txtdate), df1$arvDate, df1$txtdate)
df1$arvDate <- as.Date(df1$arvDate,format='%Y-%m-%d')

df1 <- df1[, !(colnames(df1) %in% c("newdate", "txtdate"))]
```

Range of Departure dates

```r
range(df1$depDate)
```

```
## [1] "2005-05-29" "2014-12-31"
```

Range of Arrival dates

```r
range(df1$arvDate)
```

```
## [1] "2013-01-01" "2014-12-31"
```

Arrivals before departures?

```r
table(ifelse(df1$arvDate<df1$depDate,'Arrival before departure', 'Ok'))
```

```
##
## Arrival before departure                        Ok
##                       13                    412340
```

Check the ICES rectangle and FAO areas (fields 18-19)

List of FAO Areas

```
## [1] "27.4.b" "27.4.a" "27.4.c" "27.7.d" "27.3.a"
```

Correct the areas format, if necessary.

List of corrected formated FAO areas.

```r
unique(df1$area)
```

```
## [1] "27.4.b" "27.4.a" "27.4.c" "27.7.d" "27.3.a"
```

Check the Fishing aggregation level (Metier) in each FAO area (field 20)

```r
##Read in the Reference table with the metier ib each FAO area
df2 <- readRDS(file.path(lookup.dir,'metiersAreas.rds'))

#create a table of mismatches add a column of 1s to be able to count the occurences

df3 <- df1[-which(paste(df1$foCatEu6,df1$area, sep=" ")
                  %in% paste(df2$level6,df2$codeEu, sep =" ")),]
```

A table of mismatches showing the frequency of their occurence

```r
if (is.data.frame(df3) & nrow(df3)==0) {"No mismatches"
}else{
  df3$cn<- 1
  MismatchMetieres <- ddply(df3, .(vslFlgCtry, area, foCatEu6), summarise,
                            count = sum(cn,na.rm = TRUE))
  MismatchMetieres
  }
```

```
## [1] "No mismatches"
```

Vessel Length class

Using the Vessel Length Function created, correct and standardize the the vessel length classes, following the DCF aggregation.

DCF Vessel length classes:

VL0006 = Vessel less that 6 meters in length.

VL0010 = Vessel between 0 meters and 10 meters in length.

VL0612 = Vessel between 6 meters and 12 meters in length

VL1012 = Vessel between 10 meters and 12 meters in length.

VL1218 = Vessel between 12 meters and 18 meters in length.

VL1824 = Vessel between 18 meters and 24 meters in length.

VL2440 = Vessel between 24 meters and 40 meters in length.

VL40XX = Vessel greater than 40 meters in length.

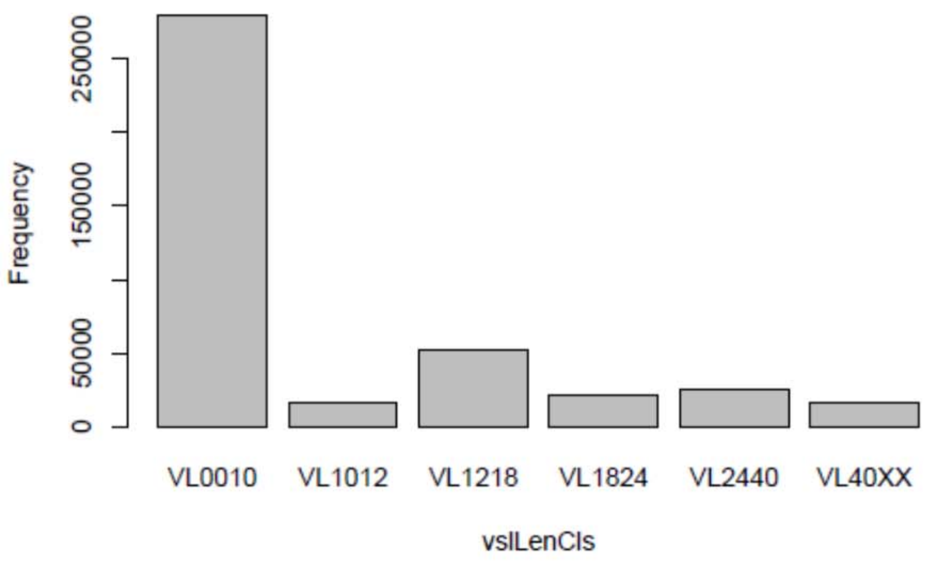
```
unique(df1$vslLenCls)
```

```
## [1] "VL1824" "VL2440" "VL0010" "VL1218" "VL1012" "VL40XX"
```

```
##Apply the Vessel length function
df1<- f.vslLenCls(df1)
```

Plot of the number of vessel length classes once corrected.

```
barplot(table(df1$vslLenCls),xlab="vslLenCls", ylab="Frequency")
```



Check Vessel type (vslType)

This code creates an additional field vslType.tmp and calculates a new vslType based on the current DCF definition and then compares it against the original data set.

```r
# Set a temporary vslType, calculates dayAtSea, and identifies dominant/polyvalent vslTyp
# Need a lookup_table metiers_TO_vslType.tmp
df1$vslType.tmp   <- RefTblDCFGears$vslType[match(substring(df1$foCatEu6,1,3),
                                             RefTblDCFGears$DCFGear)]

df1$yy <- as.numeric(substring(df1$depDate,1,4))
df1$dayAtSea <- difftime(df1$arvDate,df1$depDate,unit='days')+1

dayAtSea <-dcast(df1, yy+vslId+voyageId+vslType.tmp~., mean, value.var='dayAtSea')
dayAtSea <-dcast(dayAtSea, yy+vslId~vslType.tmp, sum, value.var='.')
tot.dayAtSea <- rowSums(dayAtSea[,3:length(dayAtSea)])
max.dayAtSea <- apply(dayAtSea[,3:length(dayAtSea)],1,max)
max.vslType <- apply(dayAtSea[,3:length(dayAtSea)],1,
                     function(x){names(dayAtSea)[which.max(x)+2]})

passive.gears <- RefTblDCFGears$vslType[RefTblDCFGears$GearType=='Passive']
passive.gears <- rowSums(dayAtSea[,which(names(dayAtSea) %in% passive.gears)])
active.gears <- RefTblDCFGears$vslType[RefTblDCFGears$GearType=='Active']
active.gears <-  rowSums(dayAtSea[,which(names(dayAtSea) %in% active.gears)])

vslType <- ifelse(max.dayAtSea>0.5*tot.dayAtSea, max.vslType,
                  ifelse(passive.gears>0 & active.gears>0,'PMP',
                         ifelse(passive.gears>0, 'PGP', 'MGP')
                  )
)

df1$vslType.tmp <- vslType[match(interaction(df1$yy,df1$vslId),
                                 interaction(dayAtSea$yy,dayAtSea$vslId))]
```

Table comparing vslType.tmp (horizontal) against vslType (vertical).

```r
with(df1,table(df1$vslType,df1$vslType.tmp))
```

```
##
##          DFN    DTS    HOK    PGO    PGP    PMP    TBB
##   DFN 170449    823    161     70     21    102      0
##   DTS   3651 185831     42      0      0    606   2642
##   HOK     63     16   2636      0    387      0      0
##   PMP    325      0      0   8299      6    831    302
##   TBB     13   1177      0      0      0      0  33900
```

```r
100*prop.table(table(ifelse(df1$vslType==df1$vslType.tmp,"%matching", "%not matching")))
```

```
##
##     %matching %not matching
##     95.463596      4.536404
```

```r
df1<- df1[, !(colnames(df1) %in% c("yy", "dayAtSea"))]
```

Duplicate rows

Number of duplicate rows in the dataset.

```
## [1] 14371
```

Vessel Power

Vessel power is checked and corrected to the ranges <=221KW and >221KW for TBB. Vessel Power only needed for TBB; the powers from the other vessel types set to NA.

```r
### check and clean vslPower
## Turn back into numbers

df1$vslPower.tmp <- if(is.numeric(df1$vslPower)){
  ifelse(df1$vslType!='TBB', 'NA',
      ifelse(df1$vslPower<222,'<=221KW','>221KW'))

} else {

  if(is.factor(df1$vslPower) | is.character(df1$vslPower)){
    ifelse(df1$vslType!='TBB','NA',
      ifelse(substring(df1$vslPower,1,4) %in% c('<221', '<=22'), '<=221KW',
        ifelse(substring(df1$vslPower,1,4)=='>221','>221KW','Correct vslPower class')))

  } else {

    'check data type'}
}
#table(df1$vslPower)
100*prop.table(table(ifelse(df1$vslPower==df1$vslPower.tmp,
                            " % matches with range", "% mismatches with range")))
```

```
##
##    % matches with range % mismatches with range
##                8.513474               91.486526
```

Table with the corrected vessel power ranges:

```r
df1$vslPower <- df1$vslPower.tmp
table(df1$vslPower.tmp)
```

```
##
## <=221KW  >221KW      NA
##   14273   20798  376876
```

```r
df1 <- df1[, !(colnames(df1) %in% c("vslPower.tmp"))]
```

Ports (Fields 10-12, 16-17)

Table provides a list of the provided port codes and their location that cannot be found in the UNLOCODE list

```r
df1$match.homePort <- match(df1$homePort, UNLOCODE$loCode, nomatch=0)
df1$match.depLoc <- match(df1$depLoc, UNLOCODE$loCode, nomatch=0)
df1$match.arvLoc<- match(df1$arvLoc, UNLOCODE$loCode, nomatch=0)
df1$match.atSeaSampLoc <- match(df1$atSeaSampLoc, UNLOCODE$loCode, nomatch=0)
df1$match.onshoreSampLoc<- match(df1$onShoreSampLoc, UNLOCODE$loCode, nomatch=0)

df1$match.homePort <- ifelse(df1$match.homePort>0, 1, 0)
df1$match.depLoc <- ifelse(df1$match.depLoc >0, 1, 0)
df1$match.arvLoc<- ifelse(df1$match.arvLoc>0, 1, 0)
df1$match.atSeaSampLoc <- ifelse(df1$match.atSeaSampLoc>0, 1, 0)
df1$match.onshoreSampLoc<- ifelse(df1$match.onshoreSampLoc>0, 1, 0)

mismatches <- subset (df1, df1$match.homePort == 0 | df1$match.depLoc==0 |
                      df1$match.arvLoc==0 |df1$atSeaSampLoc==0| df1$onShoreSampLoc==0)

d1<-data.frame(matrix(nrow=0,ncol=2))
d2 <- rbind(d1, cbind('homePort',
                      paste(unique(mismatches[mismatches$match.homePort==0,10]))))
if (is.data.frame (d2) & ncol (d2)==1) { d2[,2] <- "All OK"}
d3 <- rbind(d1, cbind('depLoc',
                      paste(unique(mismatches[mismatches$match.depLoc==0,11]))))
if (is.data.frame (d3) & ncol (d3)==1) { d3[,2] <- "All OK"}
d4 <- rbind (d1, cbind('atSeaSampLoc',
                       paste(unique(mismatches[mismatches$match.atSeaSampLoc==0,16]))))
if (is.data.frame (d4) & ncol (d4)==1) { d4[,2] <- "All OK"}
d5 <- rbind (d1, cbind('onshoreSampLoc',
                       paste(unique(mismatches[mismatches$match.onshoreSampLoc==0,17]))))
if (is.data.frame (d5) & ncol (d5)==1) { d5[,2] <- "All OK"}

d <- rbind(d2, d3, d4, d5)
colnames (d) <- c("PortType", "PortCode")
d
```

```
##          PortType PortCode
## 1        homePort     NULL
## 2        homePort    GBANY
## 3        homePort    GBCCS
## 4        homePort    GBZZZ
## 5        homePort       NA
## 6        homePort    GBDGL
## 7          depLoc    GBCCS
```

```
## 8          depLoc    NULL
## 9          depLoc   GBJER
## 10   atSeaSampLoc   GBCCS
## 11   atSeaSampLoc    NULL
## 12   atSeaSampLoc   GBJER
## 13 onshoreSampLoc   GBCCS
## 14 onshoreSampLoc      NA
```

Check the recType field (Field 1)

This code standardises the recType to "on-shore scheme","at-sea scheme" or "both" and displays any values that have not been corrected. Edit the code and repeat this routine to correct any stragglers.

```
df1$recType <-tolower(df1$recType)
df2 <- c("on-shore scheme","at-sea scheme", "both")

#Create a temporary dataset only with the recTye which do not match with:
#"on-shore scheme","at-sea scheme" or "both".
df3 <- df1[-which(df1$recType %in% df2),]

#Edit differences
df1$recType [df1$recType == "on-shore"]<- "on-shore scheme"
df1$recType [df1$recType == "off-shore"]<- "at-sea scheme"

#if there is anything in this dataset (df3) then output a table with the mismatches
if (is.data.frame(df3) & nrow(df3)==0){"No mismatches"
}else{
  df3$cn<- 1
  MismatchrecType <- ddply(df3, .(vslFlgCtry, recType),
                           summarize, count = sum(cn,na.rm = TRUE))
  MismatchrecType
  }
```

```
##   vslFlgCtry recType count
## 1        GBE    none 27863
## 2        GBI    none     1
## 3        GBN    none    10
## 4        GBR    none 20268
```

Species (Fields 13-14)

This lists the species names that have been used and the **W**orms species codes. Note any NAs - this is indicative of a missing or unallocated code.

```
#Species (13-14)
levels(as.factor(df1$sppName))
```

```
##  [1] "Aequipecten opercularis"
##  [2] "Alosa"
##  [3] "Amblyraja hyperborea"
##  [4] "Anarhichas lupus"
##  [5] "Aphanopus carbo"
##  [6] "Astacilla longicornis"
##  [7] "Atherina presbyter"
##  [8] "Balistes"
##  [9] "Belone belone"
## [10] "Brosme brosme"
## [11] "Cancer pagurus"
## [12] "Caprella septentrionalis"
## [13] "Carcinus maenas"
## [14] "Cephalopoda"
## [15] "Cerastoderma"
## [16] "Chelidonichthys"
## [17] "Chimaera monstrosa"
## [18] "Clupea harengus"
## [19] "Conger conger"
## [20] "Crangon crangon"
## [21] "Crassostrea gigas"
## [22] "Cyclopterus lumpus"
## [23] "Dicentrarchus"
## [24] "Dipturus batis"
## [25] "Dipturus oxyrinchus"
## [26] "Etmopterus spinax"
## [27] "Eutrigla"
## [28] "Gadus morhua"
## [29] "Gaidropsarus vulgaris"
## [30] "Galeorhinus galeus"
## [31] "Glyptocephalus cynoglossus"
## [32] "Hippoglossoides platessoides"
## [33] "Hippoglossus hippoglossus"
## [34] "Homarus gammarus"
## [35] "Labridae"
## [36] "Lepidorhombus whiffiagonis"
## [37] "Leucoraja naevus"
## [38] "Limanda limanda"
## [39] "Lithodes maja"
## [40] "Littorina"
## [41] "Loligo"
## [42] "Lophiidae"
## [43] "Maja squinado"
## [44] "Mallotus villosus"
## [45] "Melanogrammus aeglefinus"
## [46] "Mercenaria mercenaria"
## [47] "Merlangius merlangus"
## [48] "Merluccius merluccius"
## [49] "Microstomus kitt"
## [50] "Molva dypterygia"
```

```
## [51] "Molva molva"
## [52] "Mugilidae"
## [53] "Mullus surmuletus"
## [54] "Munida sarsi"
## [55] "Mustellus mustellus"
## [56] "Mustelus asterias"
## [57] "Mytilus edulis"
## [58] "Necora puber"
## [59] "Nephrops norvegicus"
## [60] "Octopodidae"
## [61] "Osmerus eperlanus"
## [62] "Ostrea edulis"
## [63] "Palaemon serratus"
## [64] "Pandalus borealis"
## [65] "Pandalus montagui"
## [66] "Pectinidae"
## [67] "Pegusa"
## [68] "Platichthys flesus"
## [69] "Pleuronectes platessa"
## [70] "Pollachius pollachius"
## [71] "Pollachius virens"
## [72] "Raja alba"
## [73] "Raja batis"
## [74] "Raja clavata"
## [75] "Raja fullonica"
## [76] "Raja microocellata"
## [77] "Raja montagui"
## [78] "Raja radiata"
## [79] "Rajella"
## [80] "Rajidae"
## [81] "Reinhardtius hippoglossoides"
## [82] "Salmo salar"
## [83] "Salmo trutta trutta"
## [84] "Sarda sarda"
## [85] "Sardina pilchardus"
## [86] "Scaphopoda"
## [87] "Scomber scombrus"
## [88] "Scophthalmus maximus"
## [89] "Scophthalmus rhombus"
## [90] "Scyliorhinidae"
## [91] "Scyliorhinus canicula"
## [92] "Scyliorhinus stellaris"
## [93] "Sebastes"
## [94] "Sepidae"
## [95] "Solea solea"
## [96] "Sparidae"
## [97] "Sparus aurata"
## [98] "Spondyliosoma cantharus"
## [99] "Sprattus sprattus"
## [100] "Squalus acanthias"
```

```
## [101] "Styela clava"
## [102] "Styelidae"
## [103] "Thunnus alalunga"
## [104] "Trachinus draco"
## [105] "Trachurus trachurus"
## [106] "Trachyscorpia cristulata cristulata"
## [107] "Trisopterus luscus"
## [108] "Trivia arctica"
## [109] "Xiphias gladius"
## [110] "Zeus faber"
```

```
unique(as.numeric(paste(df1$sppCode)))
```

```
##   [1] 125493 126178 126484 127143 127140 107254 126438 127136 127160 126437
##  [11] 138139 127149 127150 126436 148868 127023 127138 126029 105883 126440
##  [21] 127427 126417  11782 126461 105887 126822 126758 126441 126986 127139
##  [31] 105876 125541 105869 103450 223866 127146 107350 125715 127021    213
##  [41] 150636 107276 107205  11723 107253 127186 105711 127066 125546 107398
##  [51] 105820 127094 127082 126285 126445 105824 140687 125564 127137 126175
##  [61] 107552 293753 105885 126735 105913 119024 105821  11707 127141 141741
##  [71] 105814 105923 105822 126131 103929 105815 105863 137735 140480 107651
##  [81] 105693 272030 126425 126736 126421 126375 127214 105767    104 127085
##  [91] 126447 107381 217404 140658 140656 148824 126459 127144 126228 151523
## [101] 101851 105872 127026 107616 126458 138135 141919 107649 107163 183385
```

landWt (Field 15)

The landed weight is rounded to the nearest integer except where the landed weight is less than 0.5 when it is replaced by the value of 0.5. This is to distinguish zero occurences, requested by some case studies, from an actual but small landing.

```
#LandWt (15) check and correct
#if landwt greater than 0 but less than or = 0.5 then 0.5 else
df1$landWt_tmp <- ifelse (df1$landWt > 0 & df1$landWt <= 0.5, 0.5 , df1$landWt)
df1$landWt_tmp <- ifelse (df1$landWt_tmp == 0.5, 0.5, round (df1$landWt_tmp, 0))

df1$landWt <- df1$landWt_tmp
df1 <- df1[, !(colnames(df1) %in% c("landWt_tmp"))]
```

After running the code, verifying the quality check report that no issues or errors were found and the dataset is standardize, the "new" data file is created and can be combined with the datasets from the other countries and generate a dataset for each case study. ##Save the checked file

```
saveRDS(df1, file = file.path (output.dir, "CS2_CHECKED_ENG.rds"))
```

ANNEX 8 WP2 Case study simulation and estimation code

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2 – Regional sampling design for commercial fisheries

Deliverable 2.1 – Case study simulation and estimation code

Lead: MSS, Scotland

Core Team: Alastair Pout, Liz Clarke (MSS, Scotland)
Mary Christman (MCC, USA)
Ana Ribeiro Santos, Jon Elson (CEFAS, UK England)
Patrik Börjesson (SLU Sweden)
Kirsten Birch Harkannson, Marie Storr-Paulsen (DTU Aqua, Denmark)
Laurent Dubroca, (IFREMER, France)
José Castro, José Rodríguez (IEO, Spain)
Lucia Zarauz, (AZTI, Spain)
Chun Chen (IMARES, Netherlands)
Nuno Prista (IPMA, Portugal)
Julia Wischnewski (TI-SF Germany)

Simulation and estimation code for Case study 2 - On-shore design for the North Sea demersal fisheries

This document describes the simulation code used to for each design scenario in the on-shore sampling design. The estimator used was the Horvitz-Thompson estimator from the R package “survey” (Lumley 2014) with the svyDesign function being used to specify the design, the svyTotal function being used to provided estimates of the total and standard error of the landed weight, and the svyBy function being used to generate estimates by species and country domains.

The code starts with setting up the working directory, import the necessary libraries, reference tables and the original dataset submitted by each country.

```
setwd("C:\\Users\\as15\\Documents\\C6604\\Data_Storage\\ARS\\CS2_data")

##### Add the relevant libraries
library(survey)
library(reshape2)
library(RCMfunctions)

# make some functions:
lengthUnique <- function(x) {
    length(unique(x))
}
asCharacterUnique <- function(x) {
    unique(as.character(x))
}

# This function creates a sample dataset from the original dataset, based on
# the number of samples to be samples in each group:
sampleFromAGroup <- function(x, y, nsize) {
    # x defines the sampling units to be sampled y defines the groups from which
    # to sample nsize is a vector giving the sample sizes in each group, named
    # according to the unique values in y the values in y need to match the
    # names of nsize

    nGroup <- length(nsize)
    xSamp <- NULL
    for (i in 1:nGroup) {
        indx <- unique(x[which(y == names(nsize)[i]])
        samp <- sample(indx, size = nsize[i], replace = FALSE)
        if (i == 1) {
            xSamp <- samp
        } else {
            xSamp <- c(xSamp, samp)
        }
    }
    return(xSamp)
}
```

Load the dataset, combined with all countries, cleaned and standardized. And define the current number of PSUs by country. For the countries with 0 PSU sampled, we added one because all the sampling strata need to haev at least one PSU.

```
# load the data set this is called DFS
if (!exists("DFS")) {
    # load('../WP2 data/DFS_2014.rData')
    load("outputs\\DFS_2013.Rdata")
}

pop <- DFS
remove(DFS)

# Species assessed by WKNSSK
someSppFAO <- c("COD", "HAD", "HKE", "POK", "MNZ", "LEZ", "PLE", "WIT", "SOL",
    "TUR", "POL", "BLL", "GUG", "LEM", "FLE", "MUR")

# the magic numbers are the current number of PSUs by country,

magicNumbers <- c(BEL = 2, DEU = 2, DNK = 84, FRA = 44, SCT = 291, GBR = 339,
    IRL = 2, NLD = 101, NOR = 2, SWE = 80)
```

Start of the simulation code: 1 - Define the number of simulations to run 2 - Define the sampling design to run the simulations: Random, Country, Minor and Major Ports and Country and major and minor ports 3 - Define the stratrification for each sampling design 4 - Define the number of trips sampled in each port visit

```
# number of simulation to run
nsim <- 5

#Define the sampling design
#simName <-"Random"
simName <-"LandCtry"
#simName <-"PortStrata"
#simName <-"PortCtry"

pop$sppCode <- as.character(pop$sppName3)
#uniqueSppFAO <- sort(unique(pop$sppFAO))
#uniqueSppName3 <- pop$sppName3[match(uniqueSppFAO,pop$sppFAO)]
#pop$sppFAO[is.na(pop$sppFAO)] <- uniqueSppFAO[match(pop$sppName3,uniqueSppName3)]

# here we assign corrected variables back to their original name
pop$onShoreSampLocArvDate <- paste(pop$arvLoc2,pop$arvDate)
pop$voyageId <- pop$voyageId2
pop$landCtry <- pop$ctryCode2

# Define the stratification for each sampling design
if(simName=="Random")
{
# "ALL" will be a random sample of all the sites and days
onShoreSampLocGroup <- "ALL"
onShoreSampLocGroupSampSize <- c("ALL"=sum(magicNumbers))
}

if(simName=="LandCtry")
{
# landCtry will add the country as a strata
onShoreSampLocGroupSampSize <- magicNumbers
```

```
onShoreSampLocGroup <- as.character(pop$landCtry)
}

if(simName=="PortStrata")
{
#  big ports and small ports with no country
# top 95% landed weight harbours are
x <-rankCLE(pop,what="landWt",by="arvLoc2",horiz=F,rev=F,top=0.95,
           cex.names=0.7,col=2,density=30,main="Landed Weight by Port")

pop$portStrata <-rep("minor",dim(pop)[1])
pop$portStrata[pop$arvLoc2 %in% names(x)] <-"major"
onShoreSampLocGroup <- as.character(pop$portStrata)
onShoreSampLocGroupSampSize <- c("major"=800,"minor"=147)
}

# the number of vessels/trips sampled on a site day visit we set to two.
onShoreSampLocArvDateVoyageIdSampSize <- 2

#check that the number of strata is the same as the effort
names(table(onShoreSampLocGroup)) %in%  names(onShoreSampLocGroupSampSize)

pop$onShoreSampLocGroup <-onShoreSampLocGroup

# set up a domain of interest for estimation purposes
pop$metier <-pop$foCatEu6
pop$areaMetierLevel5 <- paste(pop$area,substr(pop$metier,1,7))
domainLevels <- sort(unique(pop$areaMetierLevel5))

# Here we work out the number of voyages for each of the site and day PSUs
# and assign that total to each row of the data set according to the site day PSU
# number of voyage in each market*day
NvoyClusPop <- tapply(pop$voyageId,pop$onShoreSampLocArvDate,lengthUnique)
pop$NvoyClusPop <- NvoyClusPop[match(pop$onShoreSampLocArvDate,names(NvoyClusPop))]

#Define the total number of PSU in the population, and the number of SSU
pop$NclusPop <- lengthUnique(pop$onShoreSampLocArvDate)  # number of market*days
pop$NvoyPop <- lengthUnique(pop$voyageId) # number of trips
```

We then defined the arrays to keep the simulations results in.

```
#sppList <- sort(unique(pop$sppFAO))
sppList <- sort(unique(pop$sppCode))
ctryList <- sort(unique(pop$landCtry))
strataList <- sort(unique(pop$onShoreSampLocGroup))

Nctry <-length(ctryList)
Nstrat <-length(strataList)
Nspp <- length(sppList)

simTotEst <- array(NA,dim=c(nsim,2))
simSppEst <- array(NA,dim=c(Nspp,2,nsim))
simSppSize <- array(NA,dim=c(nsim,Nspp))
simVoyageIdSize <- array(NA,dim=c(nsim))
dimnames(simSppSize)[[2]] <- sppList
dimnames(simSppEst)[[1]] <- sppList
```

```r
sampRowNames <- list()

simDomainEst <- array(NA,dim=c(length(domainLevels),2,nsim))

simCtryEst <- array(NA,dim=c(Nctry,2,nsim))
dimnames(simCtryEst)[[1]] <- ctryList
simStrataEst <- array(NA,dim=c(Nstrat,2,nsim))
dimnames(simStrataEst)[[1]] <- strataList

pop$sppFac <- factor(pop$sppCode,levels=sppList,ordered=TRUE)
#pop$sppFac <- factor(pop$sppFAO,levels=sppList,ordered=TRUE)

#number of the port*days
nMSpop <- tapply(pop$onShoreSampLocArvDate,pop$onShoreSampLocGroup,lengthUnique)

# Print the proportion of current sampling in each country.
magicNumbers/tapply(pop$onShoreSampLocArvDate,pop$ctryCode2,
                    lengthUnique)[names(magicNumbers)]*100
```

And then we run the simulation code:

```r
for (i in 1:nsim) {

    if (nsim >= 10) {
        if ((i/round(nsim/10)) - (i%/%round(nsim/10)) == 0) {
            print(paste(i, date()))
            flush.console()
        }
    } else {
        print(paste(i, date()))
        flush.console()
    }

    # sample market-days ('on shore sampling location arrival dates') from
    # market-days grouped together (e.g. by country) first we select the
    # required number of market days from the group of market days then we
    # collect all the data for those market days in the population together into
    # a sample

    selectedOnShoreSampLocArvDates <- sampleFromAGroup(pop$onShoreSampLocArvDate,
        pop$onShoreSampLocGroup, onShoreSampLocGroupSampSize)

    samp <- pop[pop$onShoreSampLocArvDate %in% selectedOnShoreSampLocArvDates,
        ]

    # now count up how many market days there are in each group of market days
    # in the sample then add to the sampled data a field for the number of
    # market days in each group in the population and a field for the number of
    # market days in each group in the sample how many market days in the sample
    nMSsamp <- tapply(samp$onShoreSampLocArvDate, samp$onShoreSampLocGroup,
        lengthUnique)
    samp$nPopOnShoreSampLocArvDate <- nMSpop[match(samp$onShoreSampLocGroup,
        names(nMSpop))]
    samp$nSampOnShoreSampLocArvDate <- nMSsamp[match(samp$onShoreSampLocGroup,
        names(nMSsamp))]
```

```
# now count up how many voyages there are in each market day in the sample -
# for each market day this is now 'the population' of voyages to sample from
# for that market day if the required sample of voyages is bigger than the
# population of voyages
nMDpop <- tapply(samp$voyageId, samp$onShoreSampLocArvDate, lengthUnique)
onShoreSampLocArvDateSampSize <- rep(onShoreSampLocArvDateVoyageIdSampSize,
    length(nMDpop))
names(onShoreSampLocArvDateSampSize) <- names(nMDpop)
onShoreSampLocArvDateSampSize[onShoreSampLocArvDateSampSize > nMDpop] <- nMDpop[onShoreSampLoc:
    nMDpop]

# print('sample trips')
selectedVoyageId <- sampleFromAGroup(samp$voyageId, samp$onShoreSampLocArvDate,
    onShoreSampLocArvDateSampSize)
samp <- samp[samp$voyageId %in% selectedVoyageId, ]

# number of trips in the sample
nMDsamp <- tapply(samp$voyageId, samp$onShoreSampLocArvDate, lengthUnique)
samp$nPopVoyageId <- nMDpop[match(samp$onShoreSampLocArvDate, names(nMDpop))]
samp$nSampVoyageId <- nMDsamp[match(samp$onShoreSampLocArvDate, names(nMDsamp))]

# number of clusters (market days in the sample)
samp$nClusSamp <- lengthUnique(samp$onShoreSampLocArvDate)
# number of voyages in each market day
nvoyClus <- tapply(samp$voyageId, samp$onShoreSampLocArvDate, lengthUnique)

samp$nvoyClus <- nvoyClus[match(samp$onShoreSampLocArvDate, names(nvoyClus))]

# The weight for the unstratified two stage sampling
samp$weight <- samp$NclusPop/samp$nClusSamp * samp$NvoyClusPop/samp$nvoyClus
# for the stratified sampling we need
samp$weightSrat <- samp$nPopOnShoreSampLocArvDate/samp$nSampOnShoreSampLocArvDate *
    samp$NvoyClusPop/samp$nvoyClus

samp$osId <- as.numeric(factor(samp$onShoreSampLocArvDate))
samp$vId <- as.numeric(factor(samp$voyageId))

svyDRand <- svydesign(id = ~onShoreSampLocArvDate + voyageId, weights = ~weight,
    data = samp)
svyDStrat <- svydesign(id = ~onShoreSampLocArvDate + voyageId, weights = ~weightSrat,
    data = samp, strata = ~onShoreSampLocGroup)
# svyD2 <- svydesign(id=~osId+vId,data=samp,fpc=~NclusPop+NvoyClusPop)

# class(samp$sppFac) <-'factor' the random sample
res <- svyby(~landWt, by = ~sppFac, FUN = svytotal, design = svyDRand)
# and the stratified sample
res <- svyby(~landWt, by = ~sppFac, FUN = svytotal, design = svyDStrat)
# res <- svyby(~landWt,by=~sppCode,FUN=svytotal,design=svyD) simTot
# <-svytotal(~landWt,design=svyDRand)
simTot <- svytotal(~landWt, design = svyDStrat)

simTotEst[i, c(1, 2)] <- c(simTot, SE(simTot))

# estimating the weight by landing country from a random and a stratified
# scheme resRandrandLW
# <-svyby(~landWt,by=~onShoreSampLocGroup,FUN=svytotal,design=svyDRand)
```

```
    resStrat <- svyby(~landWt, by = ~onShoreSampLocGroup, FUN = svytotal, design = svyDStrat)
    resCtry <- svyby(~landWt, by = ~landCtry, FUN = svytotal, design = svyDStrat)

    simStrataEst[match(resStrat[, 1], strataList), , i] <- matrix(unlist(resStrat[,
        2:3]), byrow = FALSE, ncol = 2)
    simCtryEst[match(resCtry[, 1], ctryList), , i] <- matrix(unlist(resCtry[,
        2:3]), byrow = FALSE, ncol = 2)

    simSppEst[match(res[, 1], sppList), , i] <- matrix(unlist(res[, 2:3]), byrow = FALSE,
        ncol = 2)
    simSppSize[i, ] <- tapply(samp$voyageId, factor(samp$sppFac, levels = sppList,
        ordered = T), lengthUnique)

    simVoyageIdSize[i] <- lengthUnique(samp$voyageId)

    sampRowNames[[i]] <- rownames(samp)
}

# end of the simulation loop ------------------------------- Svae the
# outputs of the simulations
simSppSize[is.na(simSppSize)] <- 0

print("end sim")
flush.console()

popTrip <- dcast(pop, voyageId ~ sppCode, value.var = "landWt", fill = 0, sum)

print("saving simulation results")
resultsObject <- list(sampRowNames = sampRowNames, simSppEst = simSppEst, simSppSize = simSppSize,,
    simVoyageIdSize = simVoyageIdSize, simCtryEst = simCtryEst, simStrataEst = simStrataEst,
    simTotEst = simTotEst, pop = pop)
dateStamp <- date()
rdaFileName <- paste(paste(getwd(), "temp", sep = "/"), paste(paste("S", simName,
    "nsim", nsim, gsub(":", "-", dateStamp), sep = " "), ".rData", sep = ""),
    sep = "/")

pdfFileName <- paste("results\\sim results ", simName, nsim, gsub(":", "-",
    dateStamp), ".pdf", sep = " ")

save(resultsObject, file = rdaFileName)
```

Populate the arrays previously produced with the suimulations results and calculate some summary statistics.

```
# reassigning a resultsObject for plots and summsry stats.
a <-resultsObject
simSppEst <-a$simSppEst
simSppSize <-a$simSppSize
pop <-a$pop
#popTrip <-a$popTrip

simTotEst <-a$simTotEst
simStrataEst <-a$simStrataEst
simCtryEst <-a$simCtryEst

sppList <- sort(unique(a$pop$sppCode))
ctryList <- sort(unique(a$pop$landCtry))
```

```
lengthUnique <- function(x){length(unique(x))}
asCharacterUnique <- function(x){unique(as.character(x))}
######################################################
# calculate summary statistics
meanSppEst <- apply(simSppEst[,1,],1,mean,na.rm=T)
sdSppEst <- apply(simSppEst[ ,1,],1,sd,na.rm=T)
sppSampSize <- colMeans(simSppSize)

meanSdSppEst <- apply(simSppEst[,2,],1,mean,na.rm=T)
ciLoSppEst <- apply(simSppEst[ ,1,],1,quantile,0.025,na.rm=T)
ciUpSppEst <- apply(simSppEst[ ,1,],1,quantile,0.975,na.rm=T)
RSEest <- sdSppEst/meanSppEst

sppPop <- tapply(pop$landWt,pop$sppFac,sum)
sppPopSize <- tapply(pop$voyageId,pop$sppFac,lengthUnique)
sppBiasEst <- 100*(meanSppEst/sppPop-1)

sppPopSD <- apply(popTrip[,sppList],2,sd)
#sppPopSD <- apply(popTrip[,2:120],2,sd)
sppPopMean <- apply(popTrip[,sppList],2,mean)

CVpopTrip <- sppPopSD/sppPopMean

ctryPop <-tapply(pop$landWt,pop$landCtry,sum)
ciLoCtryEst <- apply(simCtryEst[ ,1,],1,quantile,0.025,na.rm=T)
ciUpCtryEst <- apply(simCtryEst[ ,1,],1,quantile,0.975,na.rm=T)

strataPop <-tapply(pop$landWt,pop$onShoreSampLocGroup,sum)
ciLoStrataEst <- apply(simStrataEst[ ,1,],1,quantile,0.025,na.rm=T)
ciUpStrataEst <- apply(simStrataEst[ ,1,],1,quantile,0.975,na.rm=T)

meanCtryEst <-apply(simCtryEst[,1,],1,mean,na.rm=T)
sdCtryEst <-apply(simCtryEst[ ,1,],1,sd,na.rm=T)
ctryRSEest <- sdCtryEst/meanCtryEst

meanTotEst <-mean(simTotEst[,1])
sdTotEst <-sd(simTotEst[,1])
totRSEest <- sdTotEst/meanTotEst

# mean RSE over fish species and country
mean(RSEest, na.rm=T)
mean(ctryRSEest, na.rm=T)
100*(mean((simTotEst[,1]-sum(pop$landWt))/sum(pop$landWt))) # overall bias
```

Plot the results for each sampling design.

```
pdf(file=pdfFileName)

plot(sppPopSize,sppSampSize,type="n")
text(sppPopSize,sppSampSize,names(sppPopSize),cex=0.7)

# histogrames of simulation distributions
# for spp landed weights
fishOfInterest <-whatFish(someSppFAO)$Scientific[-c(3,7,13,16)]

for(j in 1:length(fishOfInterest))
```

```
{
hist(simSppEst[fishOfInterest[j],1,],nclass=30,xlab="Estimated total",
     main=whatFish(fishOfInterest[j])$English,density=30,col=2)
  abline(v=sppPop[fishOfInterest[j]],col=3,lty=1,lwd=4)
  abline(v=mean(simSppEst[fishOfInterest[j],1,]),col=2,lty=1,lwd=2)
  abline(v=ciUpSppEst[fishOfInterest[j]],col=2,lty=2)
abline(v=ciLoSppEst[fishOfInterest[j]],col=2,lty=2)
}
#--------------bias and variance plots Vs pop weight----------------------

plot(sppPop,meanSppEst,ylim=c(0,max(ciUpSppEst,na.rm=TRUE)),type="n")
points(sppPop,meanSppEst,pch=16,cex=0.7)
#text(sppPop,meanSppEst,names(sppPop),cex=0.7,pos=rep(c(2,4)10))
text(sppPop,meanSppEst,whatFish(names(sppPop))$Code,cex=0.7,pos=4)
segments(sppPop,ciLoSppEst,sppPop,ciUpSppEst)
abline(0,1,col="grey")
#------------------------------------------------------------------------
#-----------bias plot------------------------------
plot(sppSampSize,sppBiasEst,type="n",ylim=c(-20,20))
points(sppSampSize,sppBiasEst,pch=16,cex=0.7)
#text(sppSampSize,sppBiasEst,names(sppSampSize),cex=0.7)
text(sppSampSize,sppBiasEst,whatFish(names(sppSampSize))$Code,cex=0.7,pos=1)
abline(h=c(-10,-5,5,10),col="grey")
#----------------------------------------------------------
#--------------bias and var plot v pop weight----------------------

plot(ctryPop,meanCtryEst,ylim=c(0,max(ciUpCtryEst,na.rm=TRUE)),type="n")
points(ctryPop,meanCtryEst,pch=16,cex=0.7)
#text(sppPop,meanSppEst,names(sppPop),cex=0.7,pos=rep(c(2,4)10))
text(ctryPop,meanCtryEst,names(ctryPop),cex=0.7,pos=4)
segments(ctryPop,ciLoCtryEst,ctryPop,ciUpCtryEst)
abline(0,1,col="grey")
#------------------------------------------------------------------------

plot(sppPop,meanSppEst,xlim=c(0,1e6),ylim=c(0,1e6),type="n")
points(sppPop,meanSppEst,pch=16,cex=0.7)
text(sppPop,meanSppEst,names(sppPop),cex=0.7,pos=2)
segments(sppPop,ciLoSppEst,sppPop,ciUpSppEst)
abline(0,1,col="grey")

plot(sppPop,meanSppEst,xlim=c(0,1e3),ylim=c(0,4e3),type="n")
points(sppPop,meanSppEst,pch=16,cex=0.7)
text(sppPop,meanSppEst,names(sppPop),cex=0.7,pos=2)
segments(sppPop,ciLoSppEst,sppPop,ciUpSppEst)
abline(0,1,col="grey")

plot(sppSampSize,sppBiasEst,type="n")
text(sppSampSize,sppBiasEst,names(sppSampSize),cex=0.7)

plot(sppSampSize,sppBiasEst,type="n",xlim=c(0,10))
text(sppSampSize,sppBiasEst,names(sppSampSize),cex=0.7)

plot(CVpopTrip,RSEest,type="n")
text(CVpopTrip,RSEest,names(CVpopTrip),cex=0.7)

plot(sppSampSize,RSEest,type="n",xlim=c(0,10))
```

```
text(sppSampSize,RSEest,names(sppSampSize),cex=0.7)

plot(sdSppEst,meanSdSppEst,type="n");abline(0,1)
text(sdSppEst,meanSdSppEst,names(sdSppEst),cex=0.7)

dev.off()
```

ANNEX 9 WP2 csPi exchange format

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2 – Regional sampling design for commercial fisheries

Deliverable 2.2 – csPi exchange format

Lead: MSS, Scotland

Core Team: Alastair Pout, Liz Clarke (MSS, Scotland)
Kirsten Birch Harkannson, (DTU Aqua, Denmark)
Laurent Dubroca, (IFREMER, France)

csPi Data Exchange Format

Reproduced here is the csPI data exchange format needed for storing sampling data collected under stratified design based sampling schemes using multi stage probability based selection methods. This format is a development of the data exchange format used in the COST project (COST 2006), and for the Regional Database (RDB). It was further developed and used during the WKRDB 2014-01 (ICES 2014) workshop in October 2014 in Aberdeen, UK, and the WKRDB 2015-01 workshop in Sète, France. For development and testing during the project the format was termed “csPi”. It should be noted that this format will evolve as more nations adopt probability based selection methods and estimation routines.

Deleted	Key	Order	Name	R-Object name	Type	Req.	Basic checks	Description	'Under construction' comments
		1	Record type	recType	String	M		Fixed value SE.	Not a key variable, but used to identify the tables type
Y			Sampling type	sampType	String	M	Code list	"S" = sea sampling, "M" = market sampling of known fishing trips, "D" = market sampling of mixed trips, "V" = vendor. 1	Deleted - This one have nothing to do with the sampling event - e.g. you could sample both M and D on a market day.
	Y	2	Sample event code	seCode	Sting	M		A unique and short code identifying the sampling event. National code.	Do we need the psuId? That is a very long key and it may not be unique in the present form.
		3	Data provider	dataProv	String	M	Code list	A means of recording who provided the data to the RDB, this may be different from the institute or body that actually did the sampling	
		4	Sampling country	sampCtry	String	M	Code list	ISO 3166−1 alpha-3 codes. The country that did the sampling	Old variable moved from TR
		5	Sampling institution	sampInst		M		The sampling institution, i.e. those who actually did the sampling. This might not be the organization who ultimately provide the data to the RDB.	
		6	Sampling method	sampMeth	String	M	Code list	"Observer" or "SelfSampling".	Old variable moved from TR
		7	Sampling team	sampTeam	String	O		An identifier for the individuals or team who collected the sample. This is likely to be mainly of interest at the national level and would be an optional field for RDB inclusion. Similar fields in the ca table could provide information on who did the age reading and should be considered.	
		8	Sampling year	seYear	Integer	M		Sampling event year	
		9	Sampling date	sampDate	String	M		"YYYY-MM-DD" (ISO 8601). The date on which the sampling occurred (see general issues below).	
		10	Sampling time	sampTime	String	M		"HH:MM". The time on which the sampling occurred (see general issues below).	This should be used mainly for on-shore where distinct sampling can occur on the same day
		11	Sampling location	sampLoc	String	M		The location where the sampling is done	The loaction of the sampling will be the scheme and strata specific element of the sampling frame. For on-shore it should be one of arvLoc, landLoc, saleLoc or buyerLoc. For at-sea it will be the vessel, and the code vslId can be used to refer to the vslId specified in the tr table
		12	Sampling location type	sampLocType	String	M	Code list	This describes the type of location where the sampling occurred. Code list of "Vessel", "Port", "Market", "Factory".	Renamed from locType to sampLocType
		13	Design class	psuType	String	M	Code list	This is the type of the psu and falls within the 4 classes proposed by WKPICS 2 (ICES 2012); the vessel x time, the trip, the site x day or the site. This is the unit of which psuTotal and psuSamp are measures	We need a code for qouta sampling - and possible more
		14	Design	design				This would be a character string specifying the type of sampling design used to collect the data, for example "stratified two stage cluster". It could link specifically to the design object descriptions used in the R survey package, and would ensure that the appropriate estimation process was followed for the specific sampling data.	
		15	Population Data	popData				This would be a character string slot that allowed the specific population data from which the sample was drawn to be identified. It could be a specific link to a cpRDB object.	Moved from being a slot to SE
		16	Sampling scheme	sampScheme				The name for the sampling scheme under which the data were collected.	
		17	Sampling strata	sampStrata				The name of the sampling stratum in the particular scheme from which the actual sample was obtained. The format of this needs discussion and/or definition.	
		18	Sampling Temporal Unit	sampTemporalUnit				The period over which the sampling is organized e.g. year, season, semester, quarter. This is the temporal sampling stratum during which sampling effort would be assumed not to change. It could include seasonal periods. This temporal period is quite distinct from a time domain, such as a quarter, for which an estimate may be needed.	
		19	Sampling Temporal Identifier	sampTemporalId				The name for temporal unit to which the data belongs. E.g. "2013" or "first season 2011" or "Qtr1 2012"	
		20	Primary Sampling Unit key	psuKey				This lists the elements in the se table that exactly allows the sampling unit to be uniquely identified. Typically it would include the scheme, the stratum, the sampling temporal identifier, the sampling date and sampling location	Kirsten - I'm not sure this will give a unique sample event - more samples on a date.
		21	Primary Sampling Unit Identifier	psuId				This is the concatenation of the actual values from the fields specified by the psuKey. These are unique to the sample, and ideally map to the elements in the population dataset (cpRDB) that can be similarly classified. The psuId would be a replicated field in all the tables in the csRDB structure.	Kirsten- I'm not particularly fond of the variable. In my opinion it is redundant information and a very long code for a key variable.
		22	Primary Sampling Unit Total	psuTotal				This is the total number of possible sampling elements in the population e.g. market days over the year, vessels in the sampling frame. This would generally be obtained from the population data that is explicitly liked to the csRDB sample data. The @$popData slot was envisaged to allow this dataset to be specifically identified.	
		23	Primary Sampling Unit Sampled	psuSampled				This is the total number of samples obtained from within the sampling stratum and will correspond to the number of unique psuId values.	
		24	Primary Sampling Unit Probability	psuSampProb				This will be the probability of selecting the sampling unit, in many instances it will be the ratio of psuTotal/psuSampled. It could however be populated independently for example where an unequal probability selection method was used.	

Deleted	Key	Order	Name	R-Object name	Type	Req.	Basic checks	Description	'Under construction' comments
		1	Record type	recType	String	M		Fixed value TR.	
Y				psuId				The unique primary sampling unit identifier for the sample. Would map to the psuId in the SE table.	
	Y	2	Sample event code	seCode	Sting	M		A unique and short code identifying the sampling event. National code.	Do we need the psuId? That is a very long key and it may not be unique in the present form.
	Y	3	Year	year	Integer	M	1900-3000		
	Y	4	Project	proj	String	M	Code list	National project name. Code list is editable.	
	Y	5	Trip code	trpCode	String	M	String 50	National coding system.?3	Currently there is a problem with how a trip is defined; for an at sea sample it is a single voyage on a fishing vessel, for an onshore sample it is defined as a trip to a market. The latter is problematical when more than one vessel is sampled because it conflates the primary sampling unit (the site and day) with the secondary sampling unit (the vessel). Therefore trpCode should be a unique code that refers to a particular vessel and a particular fishing trip. A trip to a market would have a unique psuId, within which there can be samples from more than one vessel, each with its own trp-Code. The trpCode would be the code used by the sampling institution to identify the sample, and would be distinct from the voyageId which would be the official record of the vessel's trip or landing (see general issues below).
		6	Sampling type	sampType	String	M	Code list	"S" = sea sampling, "M" = market sampling of known fishing trips, "D" = market sampling of mixed trips, "V" = vendor. 1	Should this be a key varible? - It may be nice to have in all the tables
Y			Landing country	landCtry	String	M	Code list	ISO 3166-1 alpha-3 codes: the country where the vessel is landing and selling the catch.?2	In HH now
		7	Vessel flag country	vslFlgCtry	String	M	Code list	ISO 3166-1 alpha-3 codes: the flag country of the vessel. This can be different from the landing country (see description of Landing country).?2	
		8	Vessel identifier (encrypted)	vslId	Integer	O	1-999999	Encrypted vessel identifier. Id encrypted so that no-one can map the Id to the real vessel.	Should this be a key varible? - It may be nice to have in all the tables
		9	Vessel length	vslLen	Integer	O	3-160	Over-all length in metres.	
		10	Vessel length category	vslLenCat	String	M	Code list		Is the R name correct - should be the same as in CL/CE
		11	Vessel power	vslPwr	Integer	O	4-8500	Vessel power (kW)?.5	
		12	Vessel size	vslSize	Integer	O	1-2500	Gross registered tonnes (GRT)	
		13	Vessel size unit	vslSizeUnit	String	O	Code list	A new field for the vessel size unit, gross tonnage GT and gross registered tonnage GRT are used for different sizes of vessels and are not readily convertible.	Mandatory if vessel size filled in. Set up a code list
		14	Vessel type	vslType	Integer	M	Code list	1=stern trawler, 2=side trawler, 3=gillnetter, 4=other boats.	
		15	Number of sets/hauls on trip	foNum	Integer	O	1-300	Total number of hauls/sets taken during the trip. Both the stations where biological measures were taken and the stations that were not worked up should be counted here.?8	
		16	Days at sea	daysAtSea	Integer	O	1-60	In days.?9	This needs to be calculated the same way as in CE - decimal days?
Y			Sampling country	sampCtry	String	M	Code list	ISO 3166-1 alpha-3 codes. The country that did the sampling.	What about putting this in SE
Y			Sampling method	sampMeth	String	M	Code list	Observer or "SelfSampling".	What about putting this in SE
		17	Voyage identifier	voyageId	String	O		A new field for voyage identifier. This would provide a direct link to control agency data and can be the fishing trip ID or landing ID. It would provide a link between the population data and the sampling frame	How can it be a landing ID - isn't this table only about trips? Notes on confidialyti in WKRDB 2014
Y			Harbour	harbour	String	O	Code list	Landing harbour.	There are so many locations now, so it needs to be more specific.
		18	Departure location	depLoc	String	O	Code list		Always port? I guees it depends on definition. Mandatory for at-sea
		19	Depature date	depDate	String	O		"YYYY-MM-DD" (ISO 8601). The date of departure of the trip	
		20	Departure time	depTime	String	O		"HH:MM". The time of departure of the trip.	
		21	Arrival location	arvLoc	String	M			Always port? Together with sampLoc this will be the mandatory locations.
		22	Arrival date	arvDate	String	O		"YYYY-MM-DD" (ISO 8601). The date of arrival of the trip.	
		23	Arrival time	arvTime	String	O		"HH:MM". The arrival time of the trip.	
		24	Secondary sampling unit Type	ssuType	String			The type of secondary sampling unit, e.g. this could be vessels landing at a market on a particular day in the case of onshore sampling, or a fishing trip of a particular vessel in the at-sea case. Analogous to the use of psuType in the se table.	
		25	Secondary sampling unit key	ssuKey	String			Secondary sampling unit key. The variables used to identify the secondary sampling unit, typically this would be the psuId and the unique identifier for the secondary unit e.g. the vessel or the trip.	
		26	Secondary sampling unit Identifier	ssuId	String			The secondary sampling unit identifier; the actual values as specified by the ssuKey.	Do we need this. Same argument as for the psuId
		27	Secondary sampling unit total	ssuTotal	String			The total number of secondary sampling units, for example the number of vessels landing at a market on a particular day, or the total number of fishing trips made by a particular fishing vessel.	
		28	Secondary sampling unit sampled	ssuSampled	String			The number of secondary sampling units actually sampled, e.g. the total number of different vessels sampled at a market on a particular day in the case of an onshore sample. For an at-sea sample this will be the number of trips sampled from the particular vessel.	
		29	Secondary sampling unit probability	ssuSampProb	String			The sampling probability for the secondary sampling unit. i.e the ratio of ssuTotal/ssuSampled.	

HH/TR	Deleted	Key	Order	Name	R-Object name	Type	Req.	Basic checks	Description	'Under construction' comments
HH/LF			1	Record type	recType	String	M		Fixed value HH.	
HH/LF		Y	2	Sample event code	seCode	Sting	M		A unique and short code identifying the sampling event. National code.	Do we need the psuId? That is a very long key and it may not be unique in the present form.
HH/LF		Y	3	Year	year	Integer	M	1900-3000		
HH/LF		Y	4	Project	proj	String	M	Code list	National project name. Code list is editable.	
HH/LF		Y	5	Trip code	trpCode	String	M	String 50	National coding system.	
	Y		2	Sampling type	sampType	String	M	Code list	"S" = sea sampling, "M" = market sampling of known fishing trips, "D" = market sampling of mixed trips, "V" = vendor. 1	
	Y				psuId				The unique primary sampling unit identifier for the sample. Would map to the psuId in the SE table.	
	Y		4	Vessel flag country	vslFlgCtry	String	M	Code list	ISO 31661 alpha-3 codes. The flag country of the vessel. This can be different from the landing country (see description of LandingCountry).	
HH/LF			6	Aggregation level	aggLev	String	O	Code list	H=haul. T=trip.	
LF				Landing fraction	landFrac				The landed fraction. This is a means of identifying sections of the landing that can be distinguished that are not fishing operations. Typically this could relate to different parts of a landing that came from different sea areas or were caught using a different gears on the same trip. Suitable codes would need to be determined.	
HH/LF		Y	8	Station number	staNum	Integer	M		For an at-sea trip the sequential haul or set number number. For an on-shore trip this is set to 999	
HH				Fishing Operation Type	foType				The type of fishing operation recorded, e.g. could be a haul or a set, or it could be used for purse-seines fishing for Tuna around a fish aggregation device (FAD) or as a free school. In the case of an onshore sample it could be the nature of the landed fraction.	
HH				Fishing Operation Key	foKey				The means of identifying the specific fishing operations; e.g. trip and haul number, in the onshore case market day vessel and landed fraction.	What about foType?
HH		Y		Fishing Operation Identifier	foId				Fishing operation Identifier; this would be the specific identifier for the fishing operation, as specified by the foKey.	landing fraction/foType are missing. The code is too long and the information in it are reduntant
HH			9	Fishing validity	foVal	String	O	Code list	I=Invalid. V=Valid.	
HH/LF			11	Catch registration	catReg	String	M	Code list	The catch registration. An existing field that was unclear. Registers what components of the catch can be expected in the sl tables. The parts (landings/discards) of the catch are registered as "All", "Lan", "Dis", "Non". According to current definitions "All" assumes that both "Lan" and "Dis" are sampled. Do we need a category for Catch samples?	
HH/LF			12	Species registration	sppReg	String	M	Code list	The species in the catch, registered as "All", "Par", "Non".	
HH/LF			13	Date	foDate	String	M		Date. This would be defined according to the fishing operation, and the subgroup considered that it should become foDate. There might however be database issues in the changing of a field name.	How do we define the date when sampling from a trip - on-shore
HH/LF			14	Time	foTime	String	O		Starting time. "HH:MM"... in UTC.?9	How do we define the time when sampling from a trip - on-shore
HH			15	Fishing duration	foDur	Integer	O		In minutes.10	
HH			16	Pos.Start.Lat.dec.	latIni	Dec(5)	O		Shooting (start) position in decimal degrees of latitude.?11	
HH			17	Pos.Start.Lon.dec.	lonIni	Dec(5)	O		Shooting (start) position in decimal degrees of longitude?.11	
HH			18	Pos.Stop.Lat.dec.	latFin	Dec(5)	O		Hauling (stop) position in decimal degrees of latitude?.11	
HH			19	Pos.Stop.Lon.dec.	lonFin	Dec(5)	O		Hauling (stop) position in decimal degrees of longitude.?11	
HH/LF				Economical zone	ecoZone			Code list	The economic zone. An additional spatial variable that can be used to identify areas beyond the ICES divisions.	How do we define when sampling from a trip - on-shore
HH/LF			20	Area	area	String	M	Code list	Area level 3 (level 4 for Baltic, Mediterranean, and Black Seas) in the Data Collection Regulation (EC, 2008a, 2008b).	How do we define when sampling from a trip - on-shore
HH/LF			21	Statistical rectangle	rect	String	O	Code list	Area level 5 in the Data Collection Regulation (EC, 2008a, 2008b). This is the ICES statistical rectangles (e.g. 41G9) except for the Mediterranean and Black Seas, where GFCM geographical subareas (GSAs) are used.?13	How do we define when sampling from a trip - on-shore
HH/LF			22	Subpolygon	subRect	String	O	Code list	National level as defined by each country as child nodes (substratification) of the ICES rectangles. It is recommended that this is coordinated internationally, e.g. through the Regional Coordination Meetings (EC RCMs).	How do we define when sampling from a trip - on-shore
HH			23	Main fishing depth	foDep	Integer	O		Depth from surface to groundrope in metres?.5	
HH			24	Main water depth	waterDep	Integer	O		Depth from surface in metres?.14	
HH/LF			25	Fishing activity category National	foCatNat	String	O	Code list	Country specific Fishing activity category (=?métier). National level as defined by each country as child nodes (substratification) of the level-5 codes.	
HH/LF			26	Fishing activity category European lvl 5	foCatEu5	String	O	Code list	Fishing activity category (=?métier). Level 5 as defined in a hierarchic structure in the Data Collection Regulation (EC, 2008a, 2008b).	
HH/LF			27	Fishing activity category European lvl 6	foCatEu6	String	O	Code list	Fishing activity category. Level 6 as defined in a hierarchic structure in the Data Collection Regulation (EC, 2008a, 2008b). Level 6 is further specified by the Regional Coordination Meetings (EC RCMs, Council Regulation [EC] No 1543/2000) or any later authorized revision.	
HH/LF			28	Gear type	gear	String	M	Code list		
HH/LF			29	Mesh size	meshSize	Integer	O		Stretch measure.?18	
HH/LF			30	Selection device	selDev	Integer	O	Code list	Not mounted?=?0, Exit window / selection panel?=?1, grid?=?2. A selection device is defined as a square-meshed panel or window that is inserted into a towed net.	
HH/LF			31	Mesh size in selection device	meshSizeSelDev	Integer	O		In mm. The mesh size of a square-meshed panel or window shall mean the largest determinable mesh size of such a panel or window.	
LF			3	Landing country	landCtry	String	M	Code list	ISO 3166-1 alpha-3 codes	
LF				Landing location	landLoc				The landing location. A more clearly defined field for the landing harbour. One problem with having harbour in the tr table is that it assumes that a single trip has a single landing port, this may not be the case.	
LF				Landing location type	landLocType				The landing location type. This would be a descriptor of the landing location, suggested by the subgroup and could be used to distinguish ports, auctions, fish cages,factory ships etc.	
LF				Landings date	landDate				The date of the landing. Note that there would be a arrival date (in TR) and a separate landing date (in HH), with the possibility that there may be more than one landing for a single fishing trip. This is to reflect the situation found with some fishing trips which have multiple landings.	
LF				Landing time	landTime					
LF				Sales country	saleCtry				The sale location of the landed fraction of the catch.	
LF				Sales location	saleLoc				The sale country of the landed fraction of the catch.	
LF				Sales date	saleDate				The sale date of the landed fraction of the catch.	
LF				Sales time	saleTime					
LF				Buyer location	buyerLoc				The buyer's location. This is a potential sampling location that is distinct from both the landing location and the sale location.	
HH/LF				Domain 1	domain1					
HH/LF				Domain 2	domain2					
HH/LF				Fishing Operation Total	foTotal				The fishing operation total, e.g. the total number of hauls on a trip. How this would work for the landFrac needs to be tested.	This is already in TR (foNum)
HH/LF				Fishing Operation Sampled	foSampled				The number of fishing operations sampled.	
HH/LF				Fishing Operation Probability	foSampProb				The fishing operation sampling probability, this would usually be the ratio of foTotal/foSamp and the sample weight, for the fishing operation, would be the inverse of foSampProb. foTotal, foSamp and foSampProb are analogous to the sampling probability variables for the psu and ssu in the SE and TR tables.	

Deleted	Key	Order	Name	R-Object name	Type	Req.	Basic checks	Description	'Under construction' comments
		1	Record type	recType	String	M		Fixed value SL.	
	Y	2	Sample event code	seCode	Sting	M		A unique and short code identifying the sampling event. National code.	Do we need the psuId? That is a very long key and it may not be unique in the present form.
	Y	3	Year	year	Integer	M	1900-3000		
	Y	4	Project	proj	String	M	Code list	National project name. Code list is editable.	
	Y	5	Trip code	trpCode	String	M	String 50	National coding system.	
	Y	6	Station number	staNum	Integer	M		For an at-sea trip the sequential haul or set number number. For an on-shore trip this is set to 999	
	Y	7	Fishing operation Identifier	foId				Fishing operation Identifier; this would be the specific identifier for the fishing operation, as specified by the foKey.	landing fraction/foType are missing. The code is too long and the information in it are reduntant
Y		2	Sampling type	sampType	String	M	Code list	"S" = sea sampling, "M" = market sampling of known fishing trips, "D" = market sampling of mixed trips, "V" = vendor. 1	
Y				psuId				The unique primary sampling unit identifier for the sample. Would map to the psuId in the SE table.	
Y		3	Landing country	landCtry	String	M	Code list	ISO 3166-1 alpha-3 codes.	
Y		4	Vessel flag country	vslFlgCtry	String	M	Code list	ISO 3166-1 alpha-3 codes. The flag country of the vessel. This can be different from the landing country (see description of LandingCountry).	
		8	Commercial species	commSpp				The commercial species. These are how the commercial data are likely to be recorded officially - these will include landings where species are not commercially identified but landed collectively and in some case might not be correct (Anglerfish, megrim, gurnards, sole, skates and rays). This would allow the observer(s) to identify the species within a commercial species mix when sampling.	
	Y	9	Species	spp	String	M	Code list	Scientific name in Latin (Genus species). This is to be the recognised name that corresponds to the WoRMS alphi code. (A suffix to the latin name is used for species for which Stock cannot be defined by area)	
	Y	10	Catch category	catchCat	String	M	Code list	The fate of the catch: "Dis"=discard, "Lan"=landing.	
	Y	11	Landing category	landCat	String	M	Code list	The intended usage at the time of landing. This should match the same field in CL record (whether or not the fish was actually used for this or another purpose): "IND"=industry or "HUC"=human consumption.	
Y				landFrac				The landed fraction. This is a means of identifying sections of the landing that can be distinguished that are not fishing operations. Would map to the value in the HH table.	Isn't that in the foId? OBS the foType
	Y	12	Commercial size category scale	commCatScl	String	O	Code list	Commercial sorting scale code (optional for "Unsorted").	
	Y	13	Commercial size category	commCat	Integer	O	Code list	Commercial sorting category in the given scale (optional for "Unsorted"). (EC, 2006) and later amendments when scale is "EU".	
	Y	14	Subsampling category	subSampCat	String	O	Code list	Used when different fractions of the same species are subsampled at different levels. Typically used when few large specimens are taken out from the total catch before the many small fish are subsampled.	
	Y	15	Sex	sex	String	O	Code list	M=Male, F=Female, T=Transitional2 (optional for "Unsexed").	
		16	Unit type	unitType				The type of sampling unit. This would define the sampling unit at the sl level, and could be be for example a box of fish, a section of net, a basket of mixed discards	
		17	Unit Key	unitKey				The sampling unit key would be the variables that allowed the sampling unit to be uniquely identified.	
	Y	18	Unit Identifier	unitId				The sampling unit identifier, the actual value of the sampling unit as specified by the unitKey.	
		19	Weight	wt	Integer	M		Whole weight in grammes. Decimals not allowed. Weight of the corresponding stratum (Species - Catch category - size category - Sex).	
		20	Subsample weight	subSampWt	Integer	O		Whole weight in grammes. Decimals not allowed. For sea sampling: the live weight of the subsample of the corresponding stratum. For market sampling: the sample weight is the whole weight of the fish measured (e.g. the summed weight of the fish in one or more boxes).	
		21	Total weight derivation	totWtDeriv				The total weight derivation. At present there is a wt field in the SL table defined as the whole weight in grams. How that weight is obtained can in practice vary considerably and the weight derivation field seeks to identify that. Weights can be obtained from actual measures, estimates derived from observers or fishing vessel crew, official logbook records, official auction weights, or extrapolated from the weight length relationships of measured individuals.	
		22	Sampling Weight Derivation	sampWtDeriv				The sample weight derivation. As with the total weight this field seeks to clarify how the sample weight is actually derived. These can, for instance, be measured weights by observers of individual or groups of fish, measured weights by fishing vessel crew or official box weights, they can be obtained from weight length relationships of fish that are measured, but not weighed.	
		23	Measurement Type	measType				The measurement type. This would be used to describe what parameter was measured e.g. width, tail, head, folk length and is related to the measure class and measure number fields in the HL table.	
		24	species code	sppCode		m		The WoRMS alphi ID for the species name, must correspond to the name in spp	
		25	species FAO code	sppFAO		o		The three alpha code for the species from the ASFIS List of Species for Fishery Statistics Purposes	
		26	Presentation	pres				Presentation; the condition in which the sample was presented, e.g. gutted, whole, headless etc. The presentation will be related to the commercial category but can differ within a single commercial category. This is pertinent to the weight fields.	
		27	Conversion Factor Weight	convFacWt				The conversion factor used to get from the presentation to the whole weight (wt) field in the sl table.	
		28	Length code	lenCode	String	O	Code list	Class: 1?mm?=?"mm", 0.5?cm?=?"scm"; 1?cm?=?"cm"; 2.5?cm?=?25?mm", 5?cm?=?"5?cm".	
		29	Sampling Unit Total	unitTotal				The sampling unit total, e.g. the total number of boxes, baskets, etc. used to quantify the sampling unit.	
		30	Sampling Units Sampled	unitSampled				The number of sampling units sampled.	
		31	Sampling Unit Probability	unitSampProb				The sampling unit probability, this would usually be the ratio of unitTotal/unitSamp and the sample weight, for the sampling unit, would be the inverse of unitSampProb. These fields are analogous to the sampling probability variables in the SE, TR, and HH tables.	

Deleted	Key	Order	Name	R-Object name	Type	Req.	Basic checks	Description	'Under construction' comments
		1	Record type	recType	String	M		Fixed value HL.	
	Y	2	Sample event code	seCode	Sting	M		A unique and short code identifying the sampling event. National code.	Do we need the psuId? That is a very long key and it may not be unique in the present form.
	Y	3	Year	year	Integer	M	1900-3000		
	Y	4	Project	proj	String	M	Code list	National project name. Code list is editable.	
	Y	5	Trip code	trpCode	String	M	String 50	National coding system.	
	Y	6	Station number	staNum	Integer	M		For an at-sea trip the sequential haul or set number. For an on-shore trip this is set to 999	
	Y	7	Fishing operation Identifier	foId				Fishing operation Identifier; this would be the specific identifier for the fishing operation, as specified by the foKey.	landing fraction/foType are missing. The code is too long and the information in it are redundant
	Y	9	Species	spp	String	M	Code list	Scientific name in Latin (Genus species). This is to be the recognised name that corresponds to the WoRMS alphi code. (A suffix to the latin name is used for species for which Stock cannot be defined by area)	
	Y	10	Catch category	catchCat	String	M	Code list	The fate of the catch: "Dis"=discard, "Lan"=landing.	
	Y	11	Landing category	landCat	String	M	Code list	The intended usage at the time of landing. This should match the same field in CL record (whether or not the fish was actually used for this or another purpose): "IND"=industry or "HUC"=human consumption.	
	Y	12	Commercial size category scale	commCatScl	String	O	Code list	Commercial sorting scale code (optional for "Unsorted").	
	Y	13	Commercial size category	commCat	Integer	O	Code list	Commercial sorting category in the given scale (optional for "Unsorted"). (EC, 2006) and later amendments when scale is "EU".	
	Y	14	Subsampling category	subSampCat	String	O	Code list	Used when different fractions of the same species are subsampled at different levels. Typically used when few large specimens are taken out from the total catch before the many small fish are subsampled.	
	Y	15	Sex	sex	String	O	Code list	M=Male, F=Female, T=Transitional2 (optional for "Unsexed").	
	Y	18	Unit Identifier	unitId				The sampling unit identifier, the actual value of the sampling unit as specified by the unitKey.	
Y		2	Sampling type	sampType	String	M	Code list	"S" = sea sampling, "M" = market sampling of known fishing trips, "D" = market sampling of mixed trips, "V" = vendor. 1	
Y				psuId				The unique primary sampling unit identifier for the sample. Would map to the value in the SE table.	
Y		3	Landing country	landCtry	String	M	Code list	ISO 3166-1 alpha-3 codes.	
Y		4	Vessel flag country	vslFlgCtry	String	M	Code list	ISO 3166-1 alpha-3 codes. The flag country of the vessel. This may be different from the landing country (see description of LandingCountry).	
Y				landFrac				The landed fraction. This is a means of identifying sections of the landing that can be distinguished that are not fishing operations. Would map to the value in the HH table.	
	Y	16	Individual sex	indSex	String	M	Code list (sex)	The individual sex field allows the sample of a species to be split by sex after being sampled e.g. Rays, Nephrops. This allows a sex specific length distribution to be recorded. This was a point of clarity for the subgroup as to why this field was needed, it is an existing field and there would be no change to its use. Compare with the sex field in the sl table.	
	Y	17	Length class	lenCls	Integer	M		In mm. Identifier: lower bound of size class, e.g. 650 for 65?-?66?cm.	
		18	Number at length (not raised to whole catch)	lenNum	Integer	M		Length classes with zero should be excluded from the record.	
	Y		Measurement type	measType				The measurement type field might need to be in this table. This is, as with the HL table, the parameter recorded e.g. fork length, max length, shell height; see the rationale for measType in the sl table.	
	Y		Measurement class	measCls				The measurement class. This would be the measurement value taken on the sampled individuals. It could be an alternative or additional measure to the length, which can already be recorded in the HL table. The measurement class would relate to the measType field in the SL table. Examples could be tail widths of Nephrops in mm, carapace lengths, and carapace widths.	
			Measurement Number	measNum				The number of individuals at the specified measurement class. This is a new field. measCls and measNum are analogous to the existing lenCls and LenNum for the recording of length frequency distributions.	
			Conversion Factor Length	convFacLen				A new field; the conversion factor needed to get from the parameter measured to a standard length. For example the value used to get from the measure to the length class such as Nephrops tails X 3 = carapace length. This conversion factor can be length dependent therefore it has to be in the hl table with the length class.	
			species code	sppCode		M		The WoRMS alphi ID for the species name, must correspond to the name in spp	
			species FAO code	sppFAO				The three alpha code for the species from the ASFIS List of Species for Fishery Statistics Purposes	
			Fish Total	fishTotal				The total number of fish (or shellfish), e.g. the total number of fish within the sample such as a box or basket.	
			Fish Measured	fishSampled				The number of fish (or shellfish) measured, e.g. the number of fish for which length or another measure was made from within the sample.	
			Fish Sampling Probability	fishSampProb				The sampling probability for the measured fish, this would usually be the ratio of fishTotal/fishSamp and the sample weight, for the measured fish would be the inverse of fishSampProb. These fields are analogous to the sampling probability variables in the SE, TR, and HH and SL tables.	

Deleted	Key	Order	Name	R-Object name	Type	Req.	Basic checks	Description	'Under construction' comments
		1	Record type	recType	String	M		Fixed value CA.	
	Y	2	Sample event code	seCode	Sting	M		A unique and short code identifying the sampling event. National code.	Do we need the psuId? That is a very long key and it may not be unique in the present form.
	Y	3	Year	year	Integer	M	1900-3000		
	Y	4	Project	proj	String	M	Code list	National project name. Code list is editable.	
	Y	5	Trip code	trpCode	String	M	String 50	National coding system.?3	
Y			Sampling type	sampType	String	M	Code list	"S" = sea sampling, "M" = market sampling of known fishing trips, "D" = market sampling of mixed trips, "V" = vendor. 1	
Y				psuId					
Y			Landing country	landCtry	String	M	Code list	ISO 3166-1 alpha-3 codes.	
Y			Vessel flag country	vslFlgCtry	String	M	Code list	ISO 3166-1 alpha-3 codes. The flag country of the vessel. This may be different from the landing country (see description of LandingCountry).	
		6	Station number	staNum	Integer	O	1?999	For an at-sea trip the sequential haul or set number number. For an on-shore trip this is set to 999	This needs to be populated if ages are collected by haul
		7	Fishing Operation Identifier	foId					
		8	Quarter	quarter	Integer	M	Code list		
		9	Month	month	Integer	O	Code list		
		10	Species	spp	String	M	Code list	Scientific name in Latin (Genus species). This is to be the recognised name that corresponds to the WoRMS alphi code. (A suffix to the latin name is used for species for which Stock cannot be defined by area)	
		11	Sex	sex	String	O	Code list	M=Male, F=Female, T=Transitional2 (optional for "Unsexed").	
		12	Unit Identifier	unitId					
		13	Individual sex	indSex					
		14	Catch category	catchCat	String	M	Code list	The fate of the catch: discard or landing.	
		15	Landing category	landCat	String	M	Code list	The intended usage at the time of landing. This should match the same field in the LS record (whether or not the fish was actually used for this or another purpose): industry or human consumption.	
Y				landFrac					why hads this been deleated? It is needed to distinguish different landed fractions of the same trip
		16	Commercial size category scale	commCatScl	String	O	Code list	Commercial sorting scale code (optional for "Unsorted").	
		17	Commercial size category	commCat	Integer	O	Code list	Commercial sorting category in the given scale. (optional for "Unsorted").	
		18	Subsampling category	subSampCat					
		19	species code	sppCode		M		The WoRMS alphi ID for the species name, must correspond to the name in spp	
		20	species FAO code	sppFAO		o		The three alpha code for the species from the ASFIS List of Species for Fishery Statistics Purposes	
		21	Stock	stock	String	O	Code list	3	
		22	Area	area	String	M	Code list	Area level 3 (level 4 for Baltic, Mediterranean, and Black Seas) in the Data Collection Regulation (EC, 2008a, 2008b).	
		23	Statistical rectangle	rect	String	O	Code list	Area level 5 in the Data Collection Regulation (EC, 2008a, 2008b). This is the ICES statistical rectangles (e.g. 41G9) except for the Mediterranean and Black Seas where GFCM geographical subareas (GSAs) are used.	
		24	Subpolygon	subRect	String	O	Code list	National level as defined by each country as child nodes (substratification) of the ICES rectangles. It is recommended that this is coordinated internationally, e.g. through the Regional Coordination Meetings (EC RCMs).	
		25	Length class	lenCls	Integer	M		In mm. Identifier: lower bound of size class, e.g. 650 for 65?-66?cm.	
		26	Age	age	Integer	O		Estimated age.	
		27	Single fish number (id)	fishId	Integer	M		National numbering system of the individual fish. Preferably unique within the given Station and Species, but necessarily unique for the given combination of key fields above.	
		28	Length code	lenCode	Integer	M	Code list	Class: 1mm="mm", 0.5cm="scm"; 1cm="cm"; 2.5cm="25mm", 5cm="5cm".	
		29	Measurement Type	measType				The measurement type field might need to be in this table. This is, as with the HL table, the parameter recorded e.g. fork length, max length, shell height; see the rationale for measType in the sl table.	
		30	Measurement Class	measCls				The measurement class might need to be in this table, as well as the length class, to indicate the increment and size of the measurements, see the rationale for measCls in the hl table length groups.	
		31	Fish at length total	fishAtLengthTotal				The total number of fish/shellfish measured (for length).	
		32	Fish at length Sampled	fishAtlengthSampled				The total number of fish/shellfish sampled for age weight maturity.	
		33	Individual Fish Sampling Probability	individualFishSampProb				The sampling probability of the fish/shellfish sampled for age weight maturity.	
		34	Aging method	ageMeth	String	O	Code list	Methodology for estimating the age.	
		35	Age-plus-group	plusGrp	String	M	Code list	+?=?Plus group, ??=?Not plus group.6	
		36	Otolith weight	otoWt	Dec(5)	O		In grammes.	
		37	Otolith side	otoSide	String	O	Code list	The side of the fish where the otolith was taken. R?=?right, L?=?left.	
		38	Weight	indWt	Dec(1)	O		In grammes.	
		39	Maturity staging method	matMeth	String	O	Code list	Methodology for estimating the maturity stage.	
		40	Maturity scale	matScale	String	O	Code list	The maturity scale gives the range of the possible stages (values).	
		41	Maturity stage	matStage	String	O	Code list	The stage (value) in the given scale.	

fishPi format help files

Package 'fishPiFormats'

May 5, 2016

Type Package

Title fishPi project data formats

Version 0.0.5

Date 25th May 2015

Author Alastair Pout

Depends R(>= 3.0.0), methods

Maintainer <a.pout@marlab.ac.uk>

Description Data formats for the fishPi project

License GPL (>= 2)

R topics documented:

`csDataTocsPi` *Converts csData to csPi format.*

Description

This function convers data in the csData format to that in the csPi data format.

Usage

```
csDataTocsPi(csobj,seobj=NULL)
```

Arguments

csobj	an object of class csData
seobj	a data frame with which to poplate the se slot of the csPi data

1

Details

csPi objects consist of 6 nested data frames: se, tr, hh, sl, hl and ca. Five of these are present in the csData structure an can be populated using the matching field names. The new se data frame can be added as an options second argument.

Value

x A csPi object combining the old csData and new se object, as outlined above

Warning

The returned csobj is no longer in csData format.

Author(s)

Alastair Pout <a.pout@marlab.ac.uk>

Examples

```
# to do
```

`csPi` *Commercial Sampling Data format for use in the fishPi project*

Description

A data exchange format for Commercial Sampling Data for use in the the fishPi project.

Format

Formal class 'csPi' [package "fishPiFormats"] objects with 11 slots:
@classVersion: a character string
@desc: A description string
@popData: A description string
@design: A description string
@history: A description string
@se: A data frame of xx variables for recording sampling event information
@tr: A data frame of xx variables for recording fishing trip information
@hh: A data frame of xx variables for recording information at the haul, set or market fraction level.
@sl: A data frame of xx variables for recording information at species, commercial category and size class level.
@hl: A data frame of xx variables for recording length frequencies and other measures on groups of fish or shellfish.
@ca: A data frame of xx variables for recording individual measures on individual fish or shellfish.

Details

This is a data exchange format for commercial sampling data derived from FishFrame, incorporating changes during the COST project, adopted as the format for the RDB and then modified at WKRDB 2014-01. The csPi form is an R object development application for use during the fishPi project.

Author(s)

Alastair Pout, Laurent Dubroca <a.pout@marlab.ac.uk>

---

piPk *Primary key for the csPi tables*

Description

function that extracts the primary keys from a csPi tables.

Usage

```
piPk(table)
```

Arguments

table	A table from the csPi object, either se, tr, hh, sl, hl or ca.

Details

This is a data exchange format for commercial sampling data derived from FishFrame, incorporating changes during the COST project, adopted as the format for the RDB and then modified at WKRDB 2014-01. The csPi form is an R object development application for use during the fishPi project.

Value

A charcter string of the primary key

Author(s)

Laurent Dubroca <laurent.dubroca@ifremer.fr>

Index

ANNEX 10 WP2 Code lists

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2 – Regional sampling design for commercial fisheries

Deliverable 2.2 – Code lists

Lead: MSS, Scotland

Core Team: Alastair Pout, Liz Clarke (MSS, Scotland)
Kirsten Birch Harkannson, (DTU Aqua, Denmark)
Laurent Dubroca, (IFREMER, France)

Package 'fishPiCodes'

June 6, 2016

Type Package

Title fishPi project code lists

Version 1.0.3

Date 25th May 2015

Author Alastair Pout

Depends R(>= 3.0.0), methods

Maintainer <a.pout@marlab.ac.uk>

Description Data formats for the fishPi project

R topics documented:

---

areas *Areas*

Description

Area codes for the fishPi project and the RDB.

Format

A data frame with columns:
code: All area codes.
area: The FAO area.
subarea: The FAO subarea code.
division: The FAO division code.
subdivision: The FAO subdivision code.
unit: The FAO unit code.
description: A description of the area to which the code relates.

Details

The EU master data register uses FAO codes to describe fishing areas. The FAO codes follow a hierachical structure through areas, subareas, division, subdivision and unit. For example the IVa subdivision in the North Sea is "27.4.a". See http://www.fao.org/fishery/area/search/en and http://ec.europa.eu/fisheries/cfp/control/codes/index_en.htm

Author(s)

Alastair Pout <a.pout@marlab.ac.uk> and Work Package 2.2 core team

Examples

```
# loading the data
data(areas)
# looking at the first 5 rows
head(areas)
# areas in FAO area 27
areas$code[grep("27",areas$code)]
```

ASFIS *The FAO ASFIS species lists.*

Description

Data frame of the ASFIS species lists as used by the FAO.

Usage

```
data(ASFIS)
```

Format

Data frame consisting of columns:

ISSCAAP: The grouping code e.g. 32 = cod hakes haddocks, 38 = sharks rays, chimaeras;

TAXOCODE: Taxanomic code

X3A_CODE: 3 alpha code for the species e.g. MAC = mackerel, POK = saithe etc

Scientific_name: Scientific name of the species

English_name: English names for the species

French_name: French name for the species

Spanish_name: Spanish name for the species

Author: Author for the species type specamine

Family: Family to which the species belongs

Order: Order to which the species belongs

Stats_data: Species for which there is capture data held by FAO

Source

FAO as of Aug 2015

http://www.fao.org/fishery/collection/asfis/en

See Also

#whatFish, function to map names and codes.

Examples

```
data(ASFIS)
head(ASFIS)
```

ASFIS_WoRMS *The FAO ASFIS species lists and WoRMS species lists.*

Description

A look-up table matching the ASFIS species lists with the WoRMS species lists.

Usage

```
data(ASFIS_WoRMS)
```

Format

Data frame consisting of columns:

TAXOCODE: ASFIS Taxanomic code X3A_CODE: ASFIS 3 alpha code for the species e.g. MAC = mackerel, POK = saithe etc Scientific_name: ASFIS Scientific name of the species AphiaID: WoRMS AphiaID number Match.type: Status of the match between ASFIS and WoRMS ScientificName: WoRMS Scientific name AphiaID_accepted: The accepted WoRMS AphiaID Kingdom: Kingdom to which the species belongs Phylum: Phylum to which the species belongs Class: Class to which the species belongs Order: Order to which the species belongs Family: Family to which the species belongs Genus: Genus to which the species belongs Subgenus: Subgenus to which the species belongs Species: Species to which the species belongs Subspecies: Subspecies to which the species belongs

Source

WoRMS via ICES as of June 2016 http://www.fao.org/fishery/collection/asfis/en http://www.fao.org/fis

See Also

#whatFish, function to map names and codes.

Examples

```
data(ASFIS_WoRMS)
head(ASFIS_WoRMS)
```

---

metiers *Metiers*

Description

Allowable metiers for the RDB by region and fishing ground.

Format

A data frame with columns:
gear: The gear code
target: The intended target assemblage
mesh: The mesh size range in mm.
select: The presence of a selection device, either 0 for not present, or 1 for present.
meshselect: The mesh size in the selection device, in mm.
level5: The level 5 metier, this is the combination of the gear and the target assemblage.
level6: The level 6 metier, this is the combination of the gear, target assemblage, the mesh size, the selection device, and the mesh size in the selection device.

Details

Allowable metiers as circulated between RCM chairs Mar 2015.

Author(s)

Alastair Pout <a.pout@marlab.ac.uk> and Work Package 2.2 core team

Examples

```
# loading the data
data(metiers)
# looking at the first 5 rows
head(metiers)
# metiers allowable in areas 27.4 and 27.7.d
metiers$level6[metiers$IV..VIId==1]
```

---

species *Species*

Description

Species names and WoRMS codes.

Format

A data frame with columns:
spp: The scientific species name
code: The AphiaID 6 digit code used by WoRMS.

Details

The species listed here are 782 names and codes that are in the RDB as of June 2015 and are accepted by the WoRMS list. WoRMS is the World Register of Marine Species See http://www.marinespecies.org/

Author(s)

Alastair Pout <a.pout@marlab.ac.uk> and Work Package 2.2 core team

Examples

```
# loading the data
data(species)
# looking at the first 5 rows
head(species)
```

UNLOCODE *UN Location Codes*

Description

UN Location Codes v 1.7.

Format

A data frame with columns:
ctryCode: The country code
loCode: The location code, a 5 letter code the first two of which are country specific, the following 3 are the unique identifier of the location
locName: The location name.
lat: Decimal latitude of the location.
lon: Decimal longitude of the location.
fishPort: Logical variable Y and N as to whether the location is a fish port.
funcCode: 8 character string relating to the location function and the following 8 columns: port, rail, road, airport, post, multi, fixed, border.
port: Logical variable Y and N as to whether the location is a port.
rail: Logical variable Y and N as to whether the location is a rail terminal.
road: Logical variable Y and N as to whether the location is a road terminal.
airport: Logical variable Y and N as to whether the location is a airport.
post: Logical variable Y and N as to whether the location is a postal exchange office.
multi: Logical variable Y and N as to whether the location is a multi functional.
fixed: Logical variable Y and N as to whether the location is a fixed transport function e.g. oil platform.
boarder: Logical variable Y and N as to whether the location is a boarder crossing.

Details

The UN LOCODE data is the United Nations Code for Trade and Transport Locations. Full details of the data can be obtained at http://www.unece.org/cefact/locode/welcome.html.

Author(s)

Alastair Pout <a.pout@marlab.ac.uk> and Work Package 2.2 core team

Examples

```
# loading the data
data(UNLOCODE)
# looking at the first 5 rows
head(UNLOCODE)
# metiers allowable in areas 27.4 and 27.7.d
# plotting locations
plot(UNLOCODE$lon,UNLOCODE$lat,pch=".")
```

---

vslType *Vessel Types and DCF gear codes*

Description

Vessel Types used for the DCF fleet segmentation and their repective DCF gear codes.

Format

A data frame with columns:
DCFGear: The DCF gaer code i.e. the first 3 characters of the metier code
NominalGearType: The DCF gear type description.
vslType: the DCF vessel type appropriate to the gear type.
GearType: Active or passive gear type classification.

Details

The DCF fleet segmantations are based on a vessel Type classification, these vessel types are in turn based on aggregations of DCF defined gaer types. This data set enables vessel Types to be mapped from gaer types. Full details of the data can be obtained at http://datacollection.jrc.ec.europa.eu/web/dcf/worr

Author(s)

Alastair Pout <a.pout@marlab.ac.uk> and Work Package 2.2 core team

Examples

```
# loading the data
data(vslType)
# looking at the first 5 rows
head(vslType)
```

`whatFish` *Names and codes for fish species*

Description

Functions to convert the scientific name, English name or X3A codes of fish species and taxonomic groupings.

Usage

```
whatFish(x)
```

Arguments

x	a character vector of the Scientific name, the English name or the X3A code for the fish

Value

whatFish A function that returns a list of the the Scientific name, English name and the X3A code regardless of which of these it is passed.

Author(s)

Alastair Pout <a.pout@marlab.ac.uk>

Examples

```
whatFish("Gadus morhua")
whatFish("POK")
```

Index

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION IN FISHERIES DATA COLLECTION"

WP2 – Regional sampling design for commercial fisheries

WP2.3 – Case Study fisheries

Deliverable CS1 Pelagic fisheries sampling designs

Lead: DTU Aqua, Denmark

Core Team: Kirsten Håkansson, Marie Storr-Paulsen, (DTU-Aqua, Denmark)
Chun Chen, Sieto Verver, Edwin van Helmond (IMARES, Netherlands)
Alastair Pout, Liz Clarke (MSS, United Kingdom)

Data providers: ILVO (Belgium), DTU-Aqua (Denmark), TI-SF (Germany), IFREMER (France), CEFAS (England, UK), IMARES (Netherlands), MSS (Scotland UK), SLU (Sweden), MI (Ireland), (AZTI) Spain, (IPMA) Portugal.

Table of Contents

1 1 Description of pelagic fisheries in the North Sea & North Atlantic area

1.1 Introduction

This case study is focusing on four small pelagic stocks - herring (*Clupea harengus)* in the North Sea (her-nsea), the two sprat stocks (*Sprattus sprattus*) in the North Sea (spr-nsea) and sprat in Kattegat and Skagerrak (spr-kask) and mackerel (*Scomber scombrus*) in the North East Atlantic (mac-nea).

For over five centuries, small pelagic forage fish have been a key element in the commercial fisheries of the North Sea. Herring has played an especially important role in the economic and political development of Northern Europe and the landings of this stock have fluctuated over the past 60 years.

Herring was the most important forage fish fishery prior to the 1970s, when the population collapsed due to unsustainable fishing during a time of reduced productivity. The stock has recovered since the 1980s as a result of a fisheries closure followed by close monitoring and enforcement of management advice, and is currently considered to be sustainably fished, now supporting a fishery carried out mainly by Norway, Denmark, the Netherlands and Scotland (Engelhard *et al*. 2014).

In the 1960s and 1970s when the herring stock declined, industrial fisheries for other forage fish species increased. There was rapid growth in the industrial fishery for sprat, although the high landings statistics for the 1970s and 1980s is suspected to include large amounts of juvenile herring misreported as sprat at that time. Monitoring of the species composition of the sprat fishery was greatly improved in the early 1990s. Since 1991, between 100–200 kt of sprat have been landed in most years.

There are presently mainly three fishing techniques used to target forage fish (such as sprat, herring, sandeel, Norway pout, sardine and anchovy) in the North Sea. The industrial fisheries (for meal and oil) uses mid-water trawls with fine mesh nets (between 8 and 32mm) and store the catch of sandeel, Norway pout, sprat, and juvenile herring in tanks. The human consumption fishery on herring uses midwater trawls (mesh size 40-44 mm) deployed from single or paired trawlers, and either store the catch in RSW (refrigerated seawater) for human consumption (Dickey-Collas *et al.* 2013) or freezes the catch on-board into blocks of sorted fish. Furthermore, there is also a purse seine fishery on herring and sprat.

Traditionally, the fishing areas with high catches of mackerel have been in the northern North Sea, around the Shetland Islands, and off the west coast of Scotland and Ireland. The southern fishery off Spain's Cantabrian coast has also accounted for significant catches. In recent years, large mackerel catches have been taken further to the north, in Icelandic and Faroese waters and even as far north as Greenland during the summer feeding migrations, areas where almost no catches were reported prior to 2008. In 2013, catches in these areas contribute approximately to half of the total reported landings. In the north-western part of the distribution areas, mackerel are sometimes caught together with herring (ICES 2014). Mackerel is mainly exploited in a directed fishery for human consumption. This fishery tends to target bigger fish and there is evidence of discarding of smaller, less marketable fish.

1.2 Discard/Slipping

The indications are that large-scale discarding is not widespread in the directed North Sea herring fishery. A number of direct-observer surveys have been conducted on Scottish, Dutch and Norwegian pelagic trawlers, (Napier *et al*., 1999; 2002; Borges *et al*., 2008; van Helmond & Overzee, 2010a). The overall discard rate was less than 5% of the landed catch. It is likely that there are different discard rates between the specific fishing types. In freezer trawlers discarding can occur through sorting the catch and through emptying the tanks via the processing belts without sorting. For both pursers and trawlers 'poor' fish quality was a significant cause of discarding. Another reason is the processing capacity of freezer trawlers when catches are abundant (Helmond and Overzee, 2010b). The strength of year classes influences discarding behaviour, particularly of undersized fish. The influence of a strong herring year classes was apparent in the composition of discards with smaller, younger fish accounting for a high proportion of the fish discarded in 2001. Since the mid-2000s the stronger recruitment of mackerel has probably led to an increase in discarding due to mixed hauls of herring and mackerel.

Historically, the bycatch of juvenile herring in the industrial sprat fisheries has been considered high. Discard or slipping is known to occur in the industrial fishery when the species composition is not correct. Before 2015 (with the inclusion of the discard ban) the bycatch rules induce slipping as an incorrect species composition would be illegal to land. However, it has been very difficult to quantify the amount of slipping as it is considered an occasionally event that can be avoided if an observer is on-board.

1.3 Fisheries description based on fishPi data

As part of the fishPi project WP2, there was a data call requesting trip-level landings and sales information from each institute involved in the project for 2013 and 2014. For case study 1, the data call requested landings data for all trips catching one of the stocks in question from all vessels using pelagic fishing methods i.e. single and twin midwater trawls and pelagic seines. The areas of interest were Kattegat and Skagerrak (ICES divisions IIIa) for sprat, the North Sea (ICES divisions IV) for sprat and herring, and for mackerel all areas within the North East Atlantic were requested. The data used for the simulations were based on collated logbook and sales note data provided by eight countries (Belgium, Denmark, England (incl. Wales and Northern Ireland), France, Germany, Ireland, Netherlands, Portugal, Scotland, Spain and Sweden) involved in data collection under the DCF data collection regulation (199/2008).

After initial data scrutinizing the data call was rephrased to include all gears, due to the fact that in some countries a substantial part of the landings are caught with bottom trawls according to national logbooks. Countries only submitting a part of their fleet e.g. the pelagic trawlers have documented that the submitted data account for the majority of their landings of the stocks.

For the purpose of the fisheries descriptions data have been checked and corrected when necessary, but no data have been left out, see appendix 1 for more details.

Landings by country

Many European countries are exploiting some of the four pelagic stocks considered in this case study, although there is a large difference in the amount taken by countries. The largest stock and also the stock were most countries have significant part of the landings is the mackerel stock where Scotland and Ireland are having the largest part of the fishery (Table 1).

Landings by location

All countries in the case study (CS) have landings of one of the stocks in question (Table 2). In total most landings are landed in Denmark followed by the Netherlands and Norway. In 2014 the most important ports for the total landings of the four pelagic stocks considered in this case study were IJmuiden, Peterhead, Killybegs, Hanstholm and Skagen located in the Netherlands, Scotland, Ireland and Denmark (Figure 1).

However, although these ports had the main landings in tons the landings were not conducted in many trips. The ports with most landings by trips, instead of weight, show a very different pattern and these ports were mainly located in France, Spain and Denmark. This pattern indicates that the main landings in the pelagic fishery are carried out by the very large fishing industrial trawlers/ purse seiners in relatively few landings events and that a very different more small scale fishery is conducted by smaller vessels landing at other locations.

Fleet

The fleet exploring the main part of the pelagic stocks landings in weight is conducted by industrial midwater trawlers (OTM), however large amount of trips are also conducted by purse seiners (PS), hand lines and pole-lines and bottom trawlers targeting demersal fish (OTB), although the amount landed by these fleets are very small (Figure 2).

Vessel length

To get a better overview of the vessel lengths impact on the fishing and landing practice, the vessel length were groped in 6 different length groups and analysed accordingly. This indicated a very large difference in the amount landed in weight relative to vessel length. For the herring stock in the North Sea 97% of the landings were conducted by vessels above 40 meters. This pattern was nearly identical for mackerel where 91% of the mackerel were caught by this group, however, especially Spain, France and Portugal has a large fraction of vessels below 40 meters (Figure 3). The fleet catching sprat was much more diverse in respect to vessel length and landings, 58% were caught with vessels above 40 meters in the North Sea and only 11% in Kattegat – Skagerrak (Table 3).

Herring fishery in the North Sea

For the North Sea herring stock Denmark is the most important fishing nation with around 110 000 t followed by the Netherlands with close to 75 000 t (Table 1). The main part of the landings are going to the Netherlands with around 135 000 t followed by Denmark with close to 80 000 t and Germany with around 40 000 t (Table 2). In 2014 the main ports for landings of North Sea herring was the Dutch IJmuiden followed by the Danish port Skagen (Figure 4). When the ports with most landings are compared with the number of trips landed per port it is seen that some ports have many trips entering although the amount landed in tonnage was relatively small. The main part of the herring fishery was conducted in the western part of the North Sea (Figure 5).

The fishery is not distributed evenly over the year and the main part of the North Sea herring fishery is conducted in the 3rd quarter, with August as the most important month. However, numbers of trips are peaking in November indicating that many smaller vessels are having a more important fishery later in the season (Figure 6).

Sprat fishery in the North Sea, Kattegat and Skagerrak

The main fishing nation on sprat is Denmark for both the small Kattegat/ Skagerrak stock and the larger North Sea stock (Table 1). Nearly all of the landings are going into Denmark as well (Table 2). In contradiction to the herring and the mackerel fishery there is not a very large difference between the ports with the main amount of the landings in tons and the ports with the main part of the trips in the sprat fishery. This is especially true for the sprat stock in Kattegat- Skagerrak. The reason for the conjunction between landings and trips is probably because larger part of smaller vessels is contributing to the total catch (Figure 7 and 8).

The main part of the sprat catches are conducted in the eastern part of the North Sea (Figure 9). Both sprat stocks are landed mainly in the 2nd half of the year with a peak in September. This is more pronounced in the North Sea than in Kattegat- Skagerrak (Figure 10 and 11).

Mackerel fishery in the North East Atlantic

Mackerel is mainly fished by Scotland and Ireland (Table 1), but very large landings are taken by countries outside the EU as well. Norway, Iceland, Greenland and the Faroes landed 1/3 of the total landings in 2014. The most important landing countries are Norway with around 140 000 t followed by the United kingdom with close to 130 000 t and the Netherlands with around 100 000 t (Table 2).

The EU part of the North East Atlantic mackerel fishery is mainly landed in UK and Dutch ports, but also landings in Norway are significant. However, in this fishery there is a very large difference in where the main part of the landings in weight are landed and were the main part of the trips are landed. Especially ports in France and Spain have many trips, but low amounts of mackerel landings (Figure 12). This could be because the fishery is conducted as a small scale fishery, the vessels were much smaller as well, but it can also be because the fishery is a bycatch fishery were only small amounts are landed at every trip. The fishery conducted in the more northern part of the area is a targeted fishery conducted with large vessels.
The mackerel is fished from the northern part of Spain and all along the French, UK and Irish coast (Figure 13).

The main landings of mackerel in tons are in 1st and 4th quarter conducted on a relatively few numbers of trips with October as the month with highest landings (Figure 14). The numbers of trips are nearly showing the opposite trend than for the total landings where the main numbers of trips is conducted in the period from May to October.

1.4 Assessment and data requirement in present assessment

Three of the four stocks considered in this CS are assessed as a category 1 stock, indicating that there is a full analytic age-based assessment for the stocks. Only the sprat in Skagerrak/ Kattegat is assessed as a category 3 stock and therefore based on survey trends. Commercial data collected by Member States (MS) are the main input data to all stock assessment together with survey information. It is therefore also relevant to look at the assessment quality as well as the total landings within a stock when a sampling design is investigated. For mackerel it is also relevant to look at the EU share of the quota/ landings compared to the total landings as several countries outside EU have large landings as well and therefore cooperation between EU and other states are important if a relevant regional sampling design is to be conducted. The two sprat stocks and the North Sea herring stock are assessed by the ICES HAWG (herring assessment working group) and the mackerel stock is managed in WGWIDE (Working Group on Widely Distributed Stocks).

Information used in stock assessment	**Sprat in the North Sea (IV)**
Assessment method	The ICES advice is based on escapement strategy, and from 2014 including a maximum for fishing mortality to ensure sustainability. The assessment is age-based (cat. 1) using data from the landings. The advice year is from 1/7 to 30/6 the following year to ensure the data from the 1st quarter survey
Survey	Trawl- and acoustic survey (IBTS and HERAS)
Survey precision	Medium
B_{lim} and B_{pa}	B_{lim} =90 000 t. B_{pa} =142 000 t
Commercial data used in stock assessment	Landings: 1974- present Discard: is known to take place but cannot be quantified (in 2015 a

	discard ban has been enforced)
Survey data used in stock assessment	Annual maturity data from IBTS 1st quarter survey
Development in recruitment success in the last 10-years.	Increasing. 2014 year class is the largest year class observed in the stock.
Sampling methods	Control sampling and self-sampling
Total landings in 2014	140 384
EU TAC in 2014	144 000
fishPi	DK: 123 500 SE: 3 900 GE: 1500 NE 2400 (94% of the total landings)

Information used in stock assessment	**Sprat in Skagerrak/Kattegat (IIIa)**
Assessment method	The ICES advice is based only on survey trends and catches and the assessment is categorised as a cat. 3 stock.
Survey	Trawl- and acoustic survey (IBTS 1st and 3rd quarter and HERAS)
Survey precision	Medium
B_{lim} and B_{pa}	Not defined
Commercial data used in stock assessment	Landings: 1974- present Discard: is known to take place but cannot be quantified (in 2015 a discard ban has been enforced)
Survey data used in stock assessment	Only as trends
Development in recruitment success in the last 10-years.	The 2014 recruitment is estimated to be good, however before this increased year class the recruitment was for several year in the lowest 25% of the time series.
Sampling methods	Control sampling and self-sampling
Total landings in 2014	18 584
EU TAC in 2014	33 300 t (2.7 times higher than the advised TAC)
fishPi	DK:16 000, SE: 2 000 (98% of the total landings)

Information used in stock assessment	**Herring in the North Sea (IV, IIIa and VIId)**
Assessment method	The ICES advice is based on EU and Norway management strategy. The assessment is age-based (cat. 1) using data from the landings.
Survey	Trawl- and acoustic survey (IBTS, SCAI and HERAS)
Survey precision	Medium
B_{lim} and B_{pa}	B_{lim} =800 000 t. B_{pa} =1 000 000 t
Commercial data used in stock assessment	Landings: 1947- present Discard: is known to take place but is considered negligible (in 2015 a discard ban has been enforced)
Survey data used in stock assessment	Annual maturity data from HERAS survey

Development in recruitment success in the last 10-years.	Recruitment was below average between 2003-2013, but above average in 2014.
Sampling methods	Control sampling
Total landings in 2014	507 485 t
EU TAC in 2014	470 000 t
fishPi	DK: 124 423 t, UK 71 018, NL: 74 647, FR: 29 679, DE: 36767, SE: 15 583, IE: 68 (69% of the total landings)

Stock	**Mackerel in the North East Atlantic**
Assessment method	The assessment is age-based (cat. 1) using data from the landings.
Survey	3 surveys; triannual egg survey, bottom - and pelagic trawl survey (MEGS , IBTS and IESSNS)
Survey precision	Medium
B_{lim} and B_{pa}	B_{lim} =1.84 mill. t. B_{pa} =2.36 mill. t
Commercial data used in stock assessment	Landings: 2000- present, landings are considered unreliable before Discard: is has been collected since 1978 but are only provided for some areas and countries. Discard is considered negligible (in 2015 a discard ban has been enforced) Tagging data 1980-2005
Survey data used in stock assessment	
Development in recruitment success in the last 10-years.	The 2002 and 2006 year classes are the strongest year classes in the time-series. The incoming 2011 also year class appears to be above average.
Sampling methods	
Total landings in 2014	1 396 000 t In 2014 Norway, Iceland, Greenland and Faroes fished 36% of the total landings
EU TAC in 2014	613 317 t
fishPi	UK: 290 920, IE: 104 967, NL: 48 356, ES: 46 710, DK: 45 967, DE: 32 290, SE: 7 101 (total =576 311 t, 41% of the total landings)

1.5 Conclusions from fisheries descriptions

Although all countries participating in this case study have registered landings from at least one of the four stocks considered Denmark, Scotland and the Netherlands were the countries contributing to the main part of the landings.

With the herring stock in the North Sea and the 2 sprat stocks (North Sea and Kattegat- Skagerrak) EU countries are targeting the main part of the stocks. However, the mackerel stock is both fished and landed by countries inside and outside the EU.

In total 59% of the catch are landed in another country than the flag country and 16% of the catch are landed in countries outside EU, which highlights the importance of regional sampling also in countries outside EU.

Unlike the demersal fishery, the main part of the pelagic fishery is landed unsorted in factories and the landings are mainly only the target species.

A very large proportion of the fish are landed on relatively few landing events, this was especially true for North Sea herring and mackerel, however the effort pattern was different between member states as Spain and France had a higher effort and lower catches.

The main parts of the landings were conducted by vessels larger than 40 meter this is especially true for mackerel and herring while sprat in Kattegat – Skagerrak was also targeted by smaller vessels

1.6 Description of current national sampling programmes

To be able to estimate different sampling scenarios each country participating in the case study delivered numbers of conducted sampling events in 2014. The annual numbers of market samples and self-sampling are presented in Table 4. The numbers in the first column represent current existing sampling events conducted by the different institutes in Europe covering a pelagic industrial sampling. However, Spain and Portugal are conducting concurrent sampling, which imply that if one of the target species in this case study has been landed in an auction instead of an factory the fish will be sampled in another program. As the main focus in this case study is sampling of the targeted pelagic fishery the Spanish and Portuguese samples have been adjusted according to the average of the other sampling programs.

2 Simulation study

2.1 Method

This study is conducted as a simulation study with bootstrapping to compare various regional sampling designs for sampling the four pelagic stocks; herring (*Clupea harengus)* in the North Sea (her-nsea), sprat (*Sprattus sprattus*) in the North Sea (spr-nsea), sprat in Kattegat and Skagerrak (spr-kask) and mackerel (*Scomber scombrus*) in the North East Atlantic (mac-nea).

The different designs were compared on how well the total weights were estimated for all stocks, but also the estimation by domains was looked upon such as flag country, quarter and area. In this data call only total landed weight was considered and no information on age or length has been incorporated in the analysis.

Data

As part of the fishPi project WP2, there was a data call requesting trip-level landings and sales information from each institute involved in the project for 2013 and 2014. For case study 1, the data call requested landings data for all trips catching one of the stocks in question from all vessels using pelagic fishing methods i.e. single and twin midwater trawls and pelagic seines. The areas of interest were Kattegat and Skagerrak (ICES divisions IIIa) for sprat, the North Sea (ICES divisions IV) for sprat and herring, and for mackerel all areas within the North East Atlantic were requested. The data used for the simulations were based on collated logbook and sales note data provided by eight countries (Belgium, Germany, Denmark, England (incl. Wales and Northern Ireland), France, Ireland, Netherlands, Portugal, Scotland, Spain and Sweden) involved in data collection under the DCF data collection regulation (199/2008).

The original data call can be found in Annex 6.

After initial data scrutinizing the data call was rephrased to include all gears, due to the fact that in some countries a substantial part of the landings are caught with bottom trawls according to national logbooks. Countries only submitting a part of their fleet e.g. the pelagic trawlers have documented that the submitted data account for the majority of their landings of the stocks.

The target population

The main purpose with the CS 1 sampling program is to sample unsorted landings sold and processed at factories, assuming sorted landings are sold at regular auctions and therefore should be sampled in a different sampling scheme e.g. combined with the demersal sampling scheme. The reason for this distinction is mainly in the way the samples can be obtained, with unsorted landings it is relatively easy to take a small representative subsample of the landing e.g. when a landing is pumped to the factory or on-board when pumped to a tank. In the data used in this project it is not possible to directly distinguish these two types of landings, but is it assumed that only landings from trips fishing with trawl and purse seiner targeting small pelagic – or more specific targeting one of the stocks in this study will be landed unsorted and therefore only these are a part of the target population. Landings from trips targeting demersal fish or from mixed fisheries are assumed to be sorted and in general sold at auction. The same is assumed for trips using passive gears e.g. gill nets. This means that the target population is not the total landing of the four stock submitted by the 8 countries, but it covers the majority of the landings in tonnage, see Appendix 1.

Only data from 2014 have been used in the simulations. The final data set consisted of 801 vessel conducting 9913 trips. The landings can be sampled in 163 port on 4433 port*days. A lot of the vessels only fish for small pelagic in a short time period during the year and will only be a part of the sampling frame in that time period. The methods used for the data scrutinizing can be found in Appendix 1.

Sampling design

Two different types of designs were tested in this case study – on-board sampling (self-sampling) and on-shore sampling (port sampling).

The self-sampling design is a single-stage cluster sampling with vessel*day (trip) as the primary sampling unit (PSU) and random selection. It is assumed that a flag country is able to sample all vessels fishing under their flag. The sampling frame for this design is a list of vessels.

The port sampling design is a two-stage cluster sampling with port*day as PSU, trip within port*day as secondary sampling unit (SSU) and random selection at each stage. It is assumed that a country will sample a landing regardless of vessel flag country. The sampling frame for the design is at first stage a list of ports and afterwards a list of vessel.

Stratification

Country

In the self-sampling simulations where vessel*day (trip) are the PSU vessel flag country was used as country strata. In the port sampling simulations with port*day as PSU country of sampling location was used as strata.

Note: It is not possible to distinguish between England and Scotland when deducting country from sampling location, since the country code in the UNLOCODE is GB.

Vessel length groups

Vessels were grouped into two strata – below and above 40 m. In total only 106 vessels were above 40 meters and 695 vessels below 40 meters (Appendix 2).

Port groups

Ports were grouped into two strata – Large and small. Large port being the group of port where 90% of the regional landings per stock were available for sampling, see Figure 15. Out of 163 locations only 23 ended up in the 'Large' group (Appendix 3).

A detailed list of ports per country and port group can be found in Appendix 4.

Sampling effort

Sampling effort used in the main part of the simulations was based on the 2014 collective effort employed by the sampling institutions which amounted to 675 sampling events, see Table 4. These sampling events were a mix between self-sampling, targeted harbour sampling and concurrent sampling programmes.

Different effort allocation regimes tested involved keeping the present sampling effort per country; redistribute the present effort between strata proportional to the landings per strata and changing the total number of sampling events. The latter only being done in a couple of sampling designs.

In scenarios with effort equal to present sampling effort per countries, but with extra strata e.g. vessel length group, the effort within country were allocated to strata proportional to landings by strata.

A minimum sampling effort was set at two samples per strata. If a stratum ended up with a single PSU's (N), then the sampling effort was set to 0, since the R survey package were unable to handle a N at 1 and the resampling was done without replacement.

Scenarios

The manipulation of the sampling stratification and the effort allocation regimes enabled potential sampling designs (scenarios) to be tested, and evaluated.

Six scenarios (1A-1F, briefly decribed in Table 5) with differing sampling strata and/or effort allocation were evaluated in the on-board sampling design with the overall effort allocation maintained at 675 sampling events present. Scenario 1F was further explored in respect to the effect of changing the overall effort. These overall effort allocations were both increasing and decreasing total sampling effort relative to present sampling levels. Scenario 1A with simple random selection of trips without any strata was used as the baseline scenario when evaluating the designs and the scenario was run with different sample effort, for comparison with the more advanced scenarios.

Six scenarios (2A-2F) with differing sampling strata and/or effort allocation were evaluated in the port-sampling design with the overall effort allocation maintained at 675 sampling. In a single scenario – 2G – a different effort allocation were explored, which resulted in an overall increased effort. Scenario 1A was also use as baseline scenario for the port sampling scenarios.

See the result section for a detailed description of each of the scenarios. Table 5 and 6 have a short description of the scenarios.

Estimation

For each simulation, total landed tons were estimated using the Horvitz-Thompson estimator for the population total and domain of interest – domains being stock, country, area and quarter. This analysis was done using the R package "survey" (Lumley 2014) functions svydesign and svytotal. Svyby was use for domain of interest estimation.

All presented results are based on the distribution of the estimated tonnage from each simulation.

The estimated tonnage is the mean of the estimated tonnage per simulation. The standard error (SE) and the relative standard error (RSE - the standard deviation of the estimate divided by the mean) were calculated for the total and per domain of interest. The latter providing a scaled measure of the precision, and thereby means of comparing the relative performance of different sampling designs.

A 95% confidence interval around the estimated tonnage was generated based on the 2.5 and 97.5 quantiles of the simulated distribution of the estimates (Q2.5 and Q97.5).

In the summary (Appendix 5: Table 14) the mean RSE across domains (MRSE) was calculated to give an indication of the precision possible by domain. Also the design effect and the effective sample size were calculated in the comparison between scenarios.

$$Design\ effect\ (D_{eff}) = \frac{s^2_{scenario}}{s^2_{srs}}$$

$$Effective\ sample\ size = \frac{n}{D_{eff}}$$

Scenario 1A1-4 was used to calculate the sample variance for simple random sampling (s^2_{srs}) by altering the number of samples (n) to fit the scenario in question.

The relative standard deviation (RSD – also known as the coefficient of variation (CV)) was used to evaluate the effect of numbers of samples on the precision of the estimates.

Simulation settings

The resampling follows the specified sampling design and is done without replacement.

In most member states the effort is spread out during the year in a systematic way for practical reasons and to ensure appropriate coverage in time. In this case study effort has not been spread in time which means that in principle all the samples in a scenario could be from a very short timeframe – however random sampling will follow the effort in the fishery more closely than systematic sampling at regular intervals throughout the year.

Number of simulations

In all the scenarios the number of simulated samples was set to 500. The Monte Carlo error from the simulations would be improved if the number of replicates was increased, since the results showed small differences between runs of the same scenario and for most of the scenarios the estimated tons are not totally in correspondence with the true tons, which would not be the case with a higher number of replicates.

Since Monte Carlo error was evident with this number of replicates, small differences in estimates e.g. SE should be use with caution when judging the scenarios.

2.2 Results

1A. Simple random sampling of trips – No strata (Baseline scenario)

In the baseline scenario all samples were assumed collected as self-sampling by the fishermen. In the baseline scenario no stratification was applied, in effect, a random sampling without replacement from a vessels list and the effort was assumed similar as the present sampling effort with 675 sampling events. To test the effect on different effort levels four extra runs were conducted with different effort allocated to provide a baseline for the other simulations tested. 1A1 increased the sampling event to 1947 to match scenario 2G where the scenario was to make a census sampling on all large harbours and vessels. 1A2 decreased the sampling events by 50%, 1A3 doubled the number of sampling events, and 1A4, the last effort level tested, had census on all large vessels (for more details see Appendix 5, Table 1).

1B. Simple random sampling of trips – Stratified by vessel flag country

As in the former scenario all samples were assumed collected as self-sampling by the fishermen. The design is a random selection of trips stratified by the vessel flag country. This scenario with country stratification resulted in 10 strata (Appendix 5, Table 2). The present effort allocation by each Member State was maintained although presently only few countries are having a self-sampling program but rather a port sampling program. A minimum of 2 samples per strata was applied (for more details see Appendix 5, Table 2).

1C. Simple random sampling of trips – Stratified by vessel flag country, present effort but weighted according to landings

The scenario is similar to the previous but instead of maintaining the present effort distribution allocated by Member State effort was maintained as the same total effort but redistributed according to the amount landed by the flag country (Appendix 5, Table 3). In this scenario some countries would need to increase the present effort and other countries could decrease the sampling effort compared to present situation.

1D. Simple random sampling of trips – Stratified by vessel length group

In scenario 1D the trips are randomly selected and stratified according to vessel length, above and below 40 meters. The present total effort was maintained but redistributed according to the amount landed by the two vessel size groups (Appendix 5, Table 4). Countries were not used as strata in this scenario and therefore only 2 strata appear. The reason for choosing the above and below 40 meter vessel length was due to the analysis conducted in the fishery description were it was shown that in 2014 for mackerel and herring between 91% and 97% of the total landings were caught by this vessel length group. However, for the 2 sprat stocks the larger vessels had a lower coverage only accounting for 11% (Kattegat – Skagerrak) and 58% (North Sea).

1E. Simple random sampling of trips – Stratified by vessel length group and flag country

The scenario is a random selection of trips, stratified as scenario 1D according to vessel length above and below 40 meters but also to flag country. Effort allocation is distributed according to the present effort by vessel flag country. In this scenario each country is therefore having 2 strata giving 18 strata altogether as 2 countries Netherlands and Portugal is only having vessels in one of the vessels groups (Appendix 5, Table 5).

1F. Simple random sampling of trips – Stratified by vessel length group and flag country effort weighted according to landings.

The scenario is a random selection of trips stratified as scenario 1E according to vessel length above and below 40 meters and flag country but effort allocation is weighted according to the landings by flag country and vessel length group. The numbers of strata is similar to the former scenario (18), but effort allocation between countries has been shifted to match the landings by country. In this scenario different effort levels were tested as well: the present effort (Appendix 5, Table 6), half of the present effort redistributed by landings per vessel length group (1F1), twice the present effort redistributed by landings per vessel length group (1F2), census for vessels above 40 m - and effort as in scenario 1f for vessels below 40 m (1F3). The result on the effective sample size and design effect can be seen in Table 5. When a new design is proposed it is off course important to test if all relevant domains are being covered. (Figure 16), how well the country coverage was (Figure 17), coverage by area (Figure 18) and by quarter (Figure 19).

In the second set of scenarios (2A – 2G) ports and thereby countries were used as sampling locations.

2A. Simple random sampling of port days+ trip, no strata

In this scenario a random selection of port days was conducted, no strata were applied and the present effort was maintained. Within each port days (PSU) one trip was randomly selected (SSU). In the data set there are 9913 trips conducted on 4433 possible port days (Appendix 5, Table 7).

2B. Simple random sampling of port days+ trip – stratified by sampling country

As in the former scenario port +day was randomly selected as being the primary sampling unit and within a port day one random trip are selected for sampling and sampling country is strata. The present effort within a given country was maintained (Appendix 5, Table 8). In this scenario some trips, included in this study, were landed on the Faroe Island and in Norway. Therefore, these two countries were included with trips landed in their country but with no sampling effort.

2C. Simple random sampling of port days+ trip – stratified by sampling country and effort weighted according to landings in landing country

Scenario 2C is similar to 2B as the port sites are randomly selected by countries and within each port day one trip are randomly selected however, the total effort is weighted by landings per sampling country and thereby the present effort is re-distributed between countries (Appendix 5, Table 9). This indicates that Faroe Island and Norway are accounted for and the sampling events in these countries are therefore included according to the landings.

2D. Simple random sampling of port days+ trip – stratified by harbour size, effort= present redistributed by landings per big/small sampling locations

This scenario is a random selection of port days, stratified by large and small harbour sites. Large harbours are determine to be where 90% of the total landings of the four stocks contained in this case study are taken place, the remaining harbours are pooled in the small harbour group. The primary sampling unit is still port and day and effort is the present effort redistributed to 90% of the effort at the large sampling locations. Within each selected port day one trip is randomly selected (Appendix 5, Table 10).

2E. Simple random sampling of port days+ trip – stratified by harbour size and country, effort= present redistributed by landings in small/ large harbours and country (within the countries present effort)

In scenario 2E harbours are divided in large and small similar to the former 2D, but the effort allocation within a country is kept and then distributed relatively in the large and small harbour groups. As the countries outside this project (Norway and Faroe Islands) are receiving landings but are not included in the total numbers of present sampling events they have received 0 sampling events in this scenario. Within each selected port day one trip is randomly selected (Appendix 5, Table 11).

2F. Simple random sampling of port days+ trip – stratified by harbour size and country, effort= present redistributed by landings in small/ large harbours and country (weighted with the countries present landings)

Scenario 2F is similar to the former 2E as harbours are divided in large and small, but the effort allocation is weighted between countries according to landings and then distributed relatively to large or small harbours. However, unlike 2E, this scenario is given a relative weight to countries according to landings and therefore the countries outside this project (Norway and Faroe Islands) are included in the sampling events (Appendix 5, Table 12).

2G. Simple random sampling of port days+ trip – stratified by harbour size and country, effort= large harbours always sampled (census) small locations are random sampled four times a year. Within a marked day all large vessels are sampled (census) and a single random small vessel selected

In this scenario 2G there is a random selection of port days stratified by sampling country and harbours (divided in large and small similar to the former scenarios) but also in connection to the size of the vessels landing in the harbour. A large harbour is always sampled if there is a landing and big vessels are always sampled (Appendix 5, Table 13). Small vessels and sampling locations are randomly sampled with lesser effort. Port: large sampling locations are always sampled and small locations are sampled 4 times a year. Vessel length: Large vessels, above 40 meters are always sampled when encountered, small vessels are sampled once per day. As this scenario is using census sampling for large harbours and vessels the effort has been increased nearly 3 fold compared to the present effort used. Figure 20 shows the estimated total and Figure 21 to 23 shows the results per domains.

Self- sampling design 1A-F

In the first of the six different scenarios conducted on board as self-sampling, the sampling design was a true random design with no strata and the 675 presently used sampling events were distributed randomly on all 9913 trips used in this case study. A setup like this give an effective sample size of 675 and the design effect on the other calculated scenarios within this self-sampling design has been estimated according to the first random design. The next scenarios were stratifying by country using the present sampling effort by flag state showing that this stratification improved the design, but it was even further improved if the sampling events were weighted according to the landings by flag country.

This indicate that most countries already have adjusted their sampling programmes according to the landings of the national fleets but that this can be even further optimised if it is coordinated on a regional level (1B and 1C). In the pelagic fleet few very large vessels are catching the main part of the total landings and for the four stocks investigated in this case study it was shown that in 2014 for mackerel and herring between 91% and 97% of the total landings were caught in the vessel length group above 40 meter. However for the 2 sprat stocks the larger vessels had a lower coverage accounting for only 11% (Kattegat – Skagerrak) and 58% (North Sea) landings. Therefore three different scenarios were tested where the sampling designs were optimised according to the vessel length group and in scenario 1D vessels above and below 40 meters were the only strata with effort allocation according to the landings in the 2 strata. The same scenario was tested but with country and vessel length as strata –keeping the present effort allocation within a country (1E) and the third scenario 1F with vessel length and country as strata – but weighting the effort allocation according to the landings of the flag country.

The output from these scenarios shows that stratifying according to country and vessel length is by far the most effective design (Table 5) especially if the effort allocation is re-distributed relatively to the landings in the strata. The effective sample size is with the same number of samples increasing nearly 10 times from the sampling design used to day in the member states (1B compared to 1F). Different effort levels were also tested in scenario 1F1- 1F3 were the sampling level was halved, doubled and in 1F3 the sampling level was set to be census for larger vessels and same level as 1F for vessels below 40 meter. The increased effort surprisingly did not improve the design, the uncertainty (SD) decreased and the effective sample size increased but it was not obvious that the census design on larger vessels were better that just an increase in the overall sampling level (1F2 compared to 1F3).

One concern according to the design when vessel lengths were used as strata (1D-F) was the lack of coverage for the sprat stocks in the larger vessel group. Therefore the effect of this design was calculated on stock basis as well (domains) to test if the two sprat stock also would benefit from the vessel length strata design. In the vessel length based sampling design both the North Sea herring and the mackerel stocks show an improved coverage with a lower SE. However, for the sprat stocks the sampling design with large and small vessel group as strata is nearly similar to the totally random design (1D sprat all compared to 1A sprat all, Appendix 5, Table 14) indicating that little is gained for the sprat stock by vessel length stratification. However, if the vessel length and the country are used in the strata the SE is improving.

Port sampling design 2A-2G

As most member states are presently having a port sampling program as their sampling design with harbours as the sampling place, it was also tested if an on shore program could be improved through seven different scenarios. A simple random sampling of marked day was tested as the first scenario and within a marked day a random trip was selected. As was the case with the self-sampling scenarios the second scenario within this design 2B was using country as strata and keeping the present effort by country and the third 2C was redistributing the effort according to the landings by country. In the later scenario effort could therefore be reshuffled compared to the present situation if a given country were having a relatively larger or smaller share of the total landings. It only slightly improves the uncertainties if effort (Table 6) was redistributed according to the landings, indicating that the present design is relatively well balanced between countries. The harbours were divided in large and small harbours with large locations being the ones were 90% of the landings per stock were available for sampling. When this scenario was tested (2D), keeping the same total effort but distributing 90% of the effort to larger sampling locations and only 10% to the smaller sampling sites the design (SD and effective sampling size) actually became worse than in the scenario were countries were the only strata (2B). However, if large and small sampling sites were kept and country added as strata (2E) the design improved significantly, although a bit surprisingly it did not improve the design if the effort was redistributed according to the landings by country (2F) but was better with the present effort per country (2E). As a final scenario it was tested if it was possible to take account of both large and small harbour sites and large and small vessels within the different harbours (2G). In this scenario effort was increased as all large sampling location and large vessels were always sampled and small sampling locations and vessels were random selected with less effort. As there was sampled census on both large sampling locations and vessel the total effort was increased nearly three times. As the effort was much higher in this last scenario the baseline was conducted with the same effort 1A1 to have a comparison. The design improved when accounting for large sampling sites and vessels (with a low SD) and the effective sample size was 125 times higher indicating that the design improved a lot by this last design.

2.3 Discussion

These different simulations offer an opportunity to consider regional sampling designs in relation to the existing sampling programmes operated by countries within the region. As such they provide an opportunity to scope out hypothetical situations and stimulate consideration of the implications of regional sampling. Some of the assumptions underlying these simulations and the major implications of regional sampling are considered below:

Sampling locations

These simulations are based on the assumption that arrival locations used to define the site and day is a location where the landings can actually be sampled and that a targeting pelagic fishery is always landed on a factory and not in a marked or fish auction. This may not be a valid assumption in all cases, as there is no clear indication of landing type (factory/market/ fish auction ect.) in the data used in this project – and it may even be problematic to achieve this information at a national level. It has been assumed that if the fishery is a targeted pelagic fishery that the catch will all be landed in factories but in especially Spain and Portugal a large fraction of these stocks may be sold on markets and should be covered by another sampling programme. Another challenge is when landings are transferred to fish factories by lorries, then the opportunity for sampling may be limited on arrival location.

In the sampling design presented in this case study it is assumed that a sampling country sample all landings regardless of vessel flag country. If a sampling country only sample vessel from their own flag country, then a large proportion of the landings will never be available for sampling. In 2014 only 59% of the landings from a vessel flag country were landed in the country of the vessel. This would off course be a change to the present designs in many countries presently only sampling their national vessels.

Self-sampling

Relatively few countries are presently using self-sampling as a design to collect data from the commercial fishery, Denmark and the Netherlands are among the countries that are using self-sampling on a regular basis. There are several challenges with using data collected as self-sampling but also some advantages. Self- sampling data should preferable be validated from other sources to ensure the correctness of the data. This can be done by a small scale harbour sampling programme were control data can be compared to the data delivered from the industry. Also VMS tracking and cross checking with logbook data can be used to verify data. It is also important to follow the self-sampling process relatively closely if, for some reason, the fishermen stop sampling due to other conflicts between scientist and the industry it is important to be able to get samples from alternatively sources. This have in Denmark been overcome by incorporating the self-sampling in the license in some fisheries. However, there are also several advantages with self- sampling, it is a cost effective way to sample, with industrial fishery you can have the advantage to get samples haul by haul and not on a trip level that a harbour sample will most likely be and the sample will often be in a much better condition (frozen directly after the catch). If a well design self- sampling program is conducted it is a very cost effective way to conduct a sample programme.

In a self-sampling design it is not necessary for other countries to sample the flag vessel as long as the vessel can keep the sample on board until it arrives in the home harbour but the effort allocation between the countries has to be optimised and re-distributed to a more optimal coverage.

Vessel length

The simulations indicated that using the vessel length as strata highly improved the sampling design. This maybe especially true for the pelagic fleet were few trips can hold a very large share of the total landings and indicating that the main part of the landings can be covered by relatively few vessels. However, to be able to implement this design every sampling site needs a full list of all vessels above 40 meter to know if the vessel is to be sampled or not.

Large / small sampling sites

To use large and small sampling sites did not improve the sampling design a lot compared to when countries were used as strata, which was a bit surprising. Only when sampling site and country were both used as strata did it improve the design.

Reallocation of sampling effort across countries

The major finding of the on-shore design is that the present allocation of sampling effort across countries and within a country can be optimised. If the country and port design was implemented, and sampling effort were free to redistribute across countries, then Germany, England, Ireland, Scotland and to a lesser degree Spain and Portugal would need to increase their on-shore sampling. For Sweden, France and Denmark there would be a decrease in sampling commitment. However, within a country nearly all countries could gain in the sampling design if the vessel length were to be considered when the sampling schemes were conducted.

Optimal numbers of sampling events were tested by altering the numbers of samples and comparing it with the CVs of the estimate for the 3 species analysed in this SC (Figure 24)

To be solved before implementation

In this project total weight was used to estimate how well a scenario was designed. However, since three out of the four stocks in this CS is assessed with a full analytical age-based assessment it would be beneficial to include biology in the simulations study and not only total weight.

Another important improvement of the CS would be to include all pelagic stocks caught by the analyzed fleets. Most of these vessels are changing target species over the year but are landing the fish in a similar way, to factories and it would therefore be more cost effective to include all relevant stocks in the sampling program.

Table 1 Landings by stock and flag country in tons in 2014.

Flag country	her-nsea	mac-nea	spr-kask	spr-nsea	Total
DNK	109,056	41,917	22,119	139,247	312,339
SCO	41,155	213,372	-	-	254,527
NLD	74,647	46,665	-	2,444	123,756
IRL	16	102,610	-	-	102,626
ENG	25,851	44,402	-	37	70,290
FRA	29,258	17,494	-	-	46,752
ESP	-	38,062	-	-	38,062
DEU	16,018	21,895	-	-	37,913
SWE	15,630	4,280	1,525	3,872	25,307
PRT	-	3,238	-	-	3,238
BEL	27	56	-	0	83
Total	311,658	533,991	23,644	145,600	1,014,893

Table 2 Landings by landing country and stock in tons in 2014.

Landing country	her-nsea	mac-nea	spr-kask	spr-nsea	Total
DK	79,017	18,951	22,602	140,738	261,308
NL	133,061	104,397	-	4,795	242,253
NO	20,446	141,736	-	-	162,182
GB	28,413	128,511	-	37	156,961
IE	1,388	83,805	-	-	85,193
DE	41,864	1,652	-	-	43,516
ES	-	40,831	-	-	40,831
FR	3,682	10,263	-	-	13,945
SE	3,340	185	1,042	30	4,597
NA	430	2,058	-	-	2,488
FO	3	1,013	-	-	1,016
PT	-	557	-	-	557
BE	14	32	-	0	47
JE	-	0	-	-	0
Total	311,658	533,991	23,644	145,600	1,014,893

Table 3 The relative amount of landing by stock and vessel length in 2014.

Vessel length	her-nsea	mac-nea	spr-kask	Spr-nsea
VL0010	0.3	0.2	0.0	0.7
VL1012	0.3	0.3	5.8	1.3
VL1218	0.3	0.6	22.8	13.9
VL1824	0.9	1.6	44.9	14.3
VL2440	1.6	5.8	15.2	11.8
VL40XX	96.5	91.3	11.3	58.0
Total	100	100	100	100

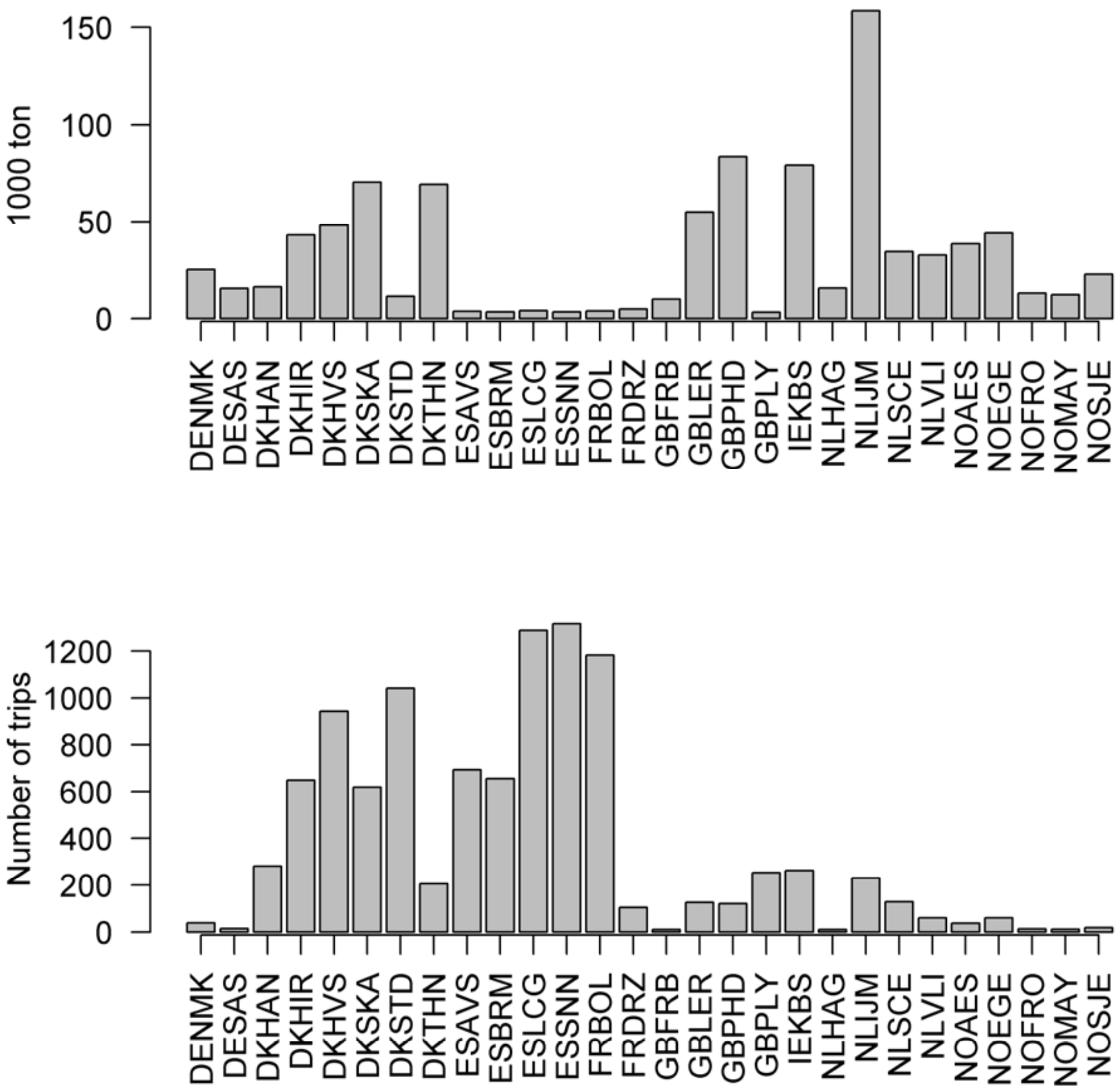


Figure 1 Total tonnage landed and trips per port in 2014. In total there were 381 ports, but the figures only display ports with at least 10 trips and 3 t of landings in total.

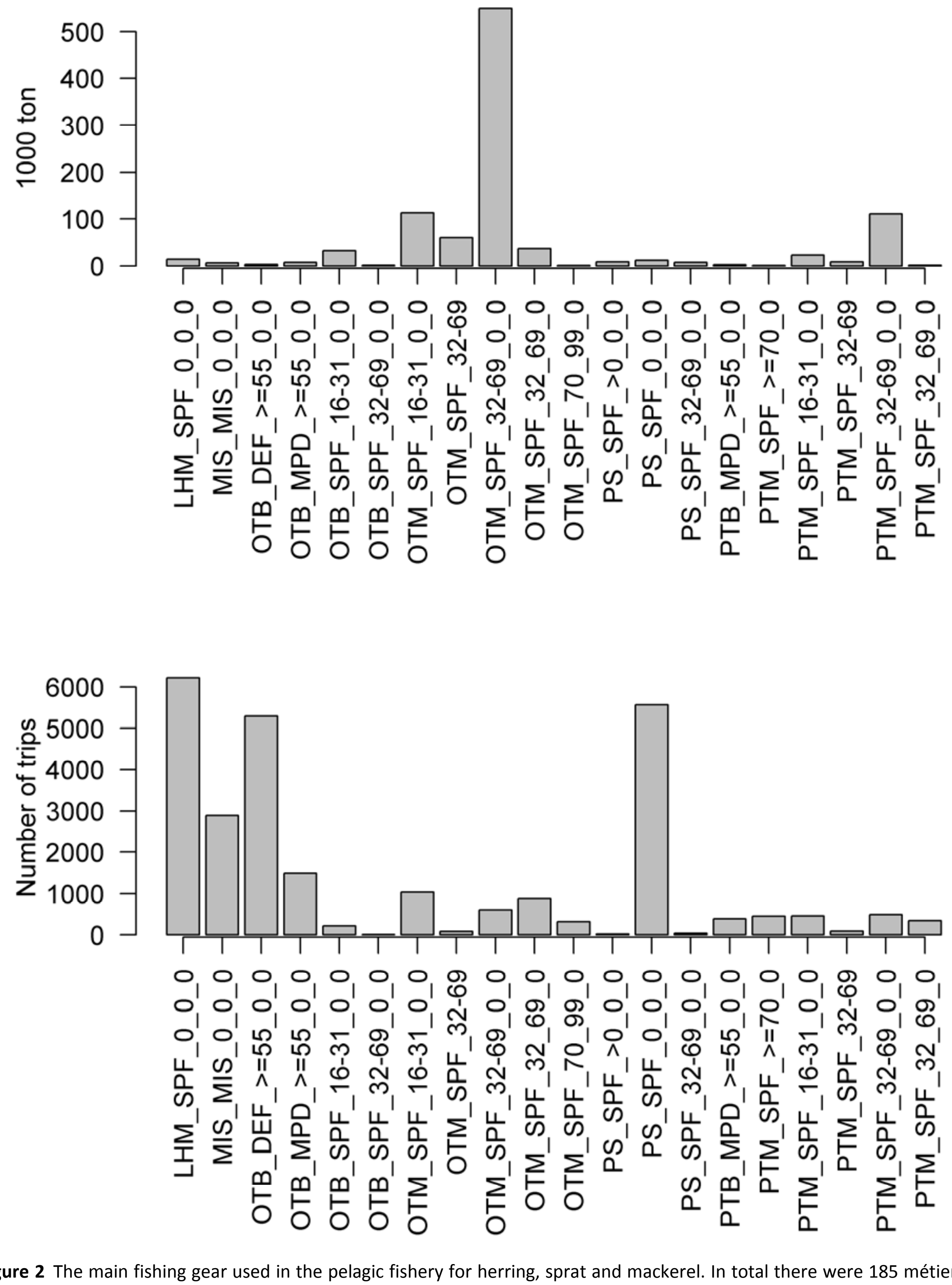


Figure 2 The main fishing gear used in the pelagic fishery for herring, sprat and mackerel. In total there were 185 métiers, but the figures only display métiers with at least 5 trips and 1 t of landings in total.

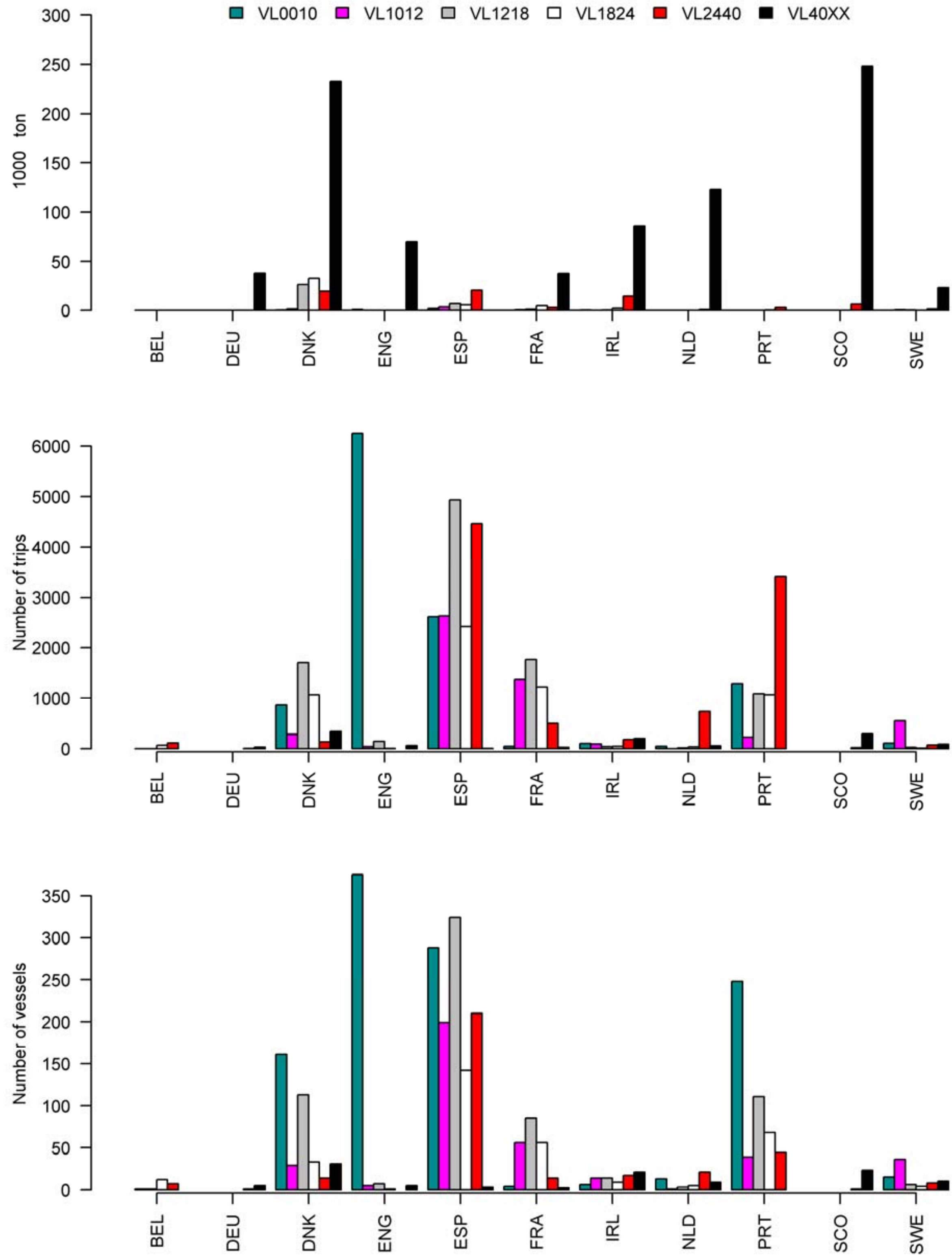


Figure 3 The total landings of the 4 pelagic stocks considered in this case study and the numbers of vessels and trips by vessel length. 6 different vessel length groups have been look at ranging from 0-10 m, 10-12 m, 12-18 m, 18-24 m, 24-40 m and above 40 m.

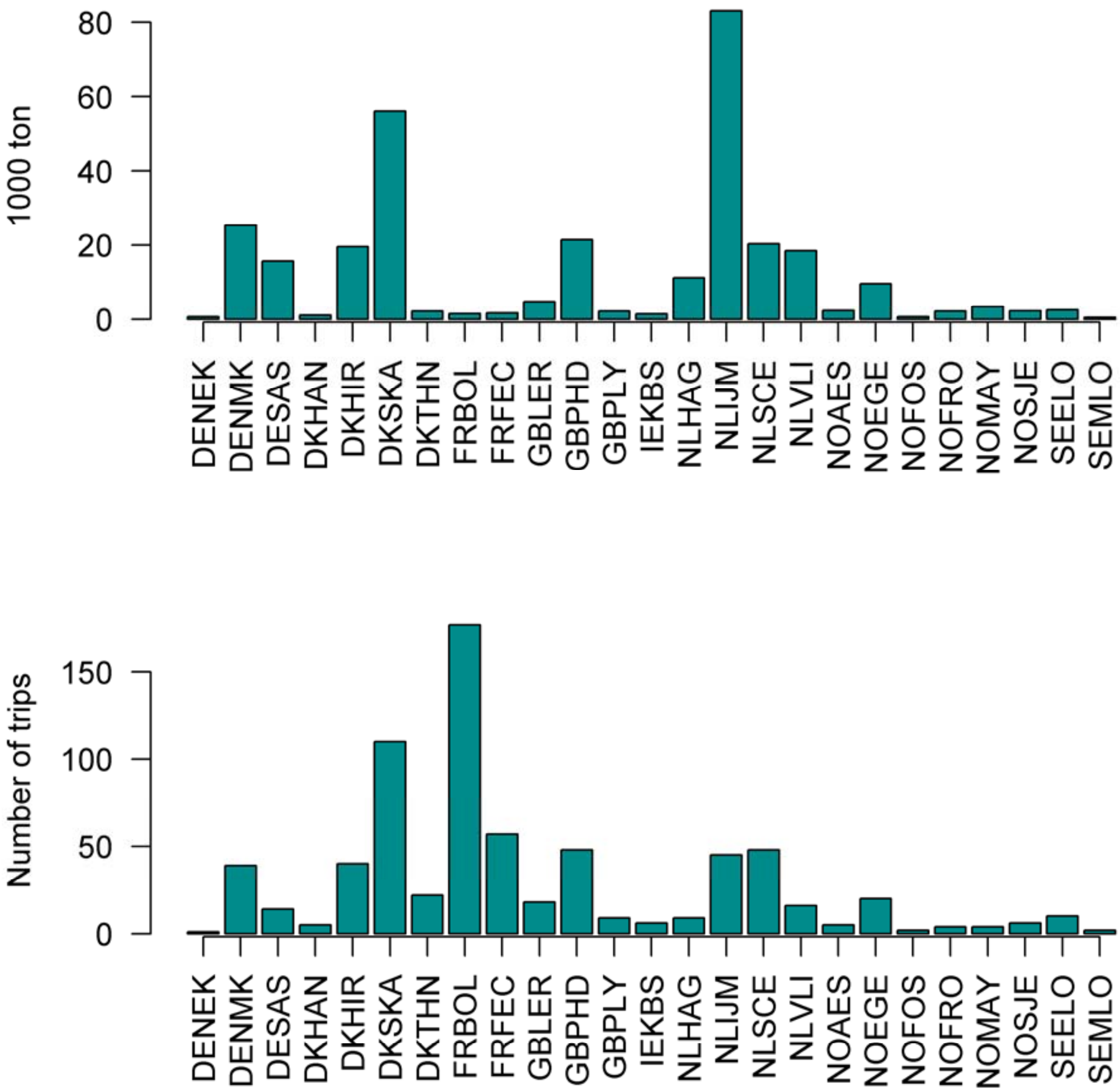


Figure 4 Tonnage landed and trips per port in 2014 for North Sea herring. There were 56 ports with landings of the stock, but the figures only display ports with at least 0.5 t of landings in 2014.

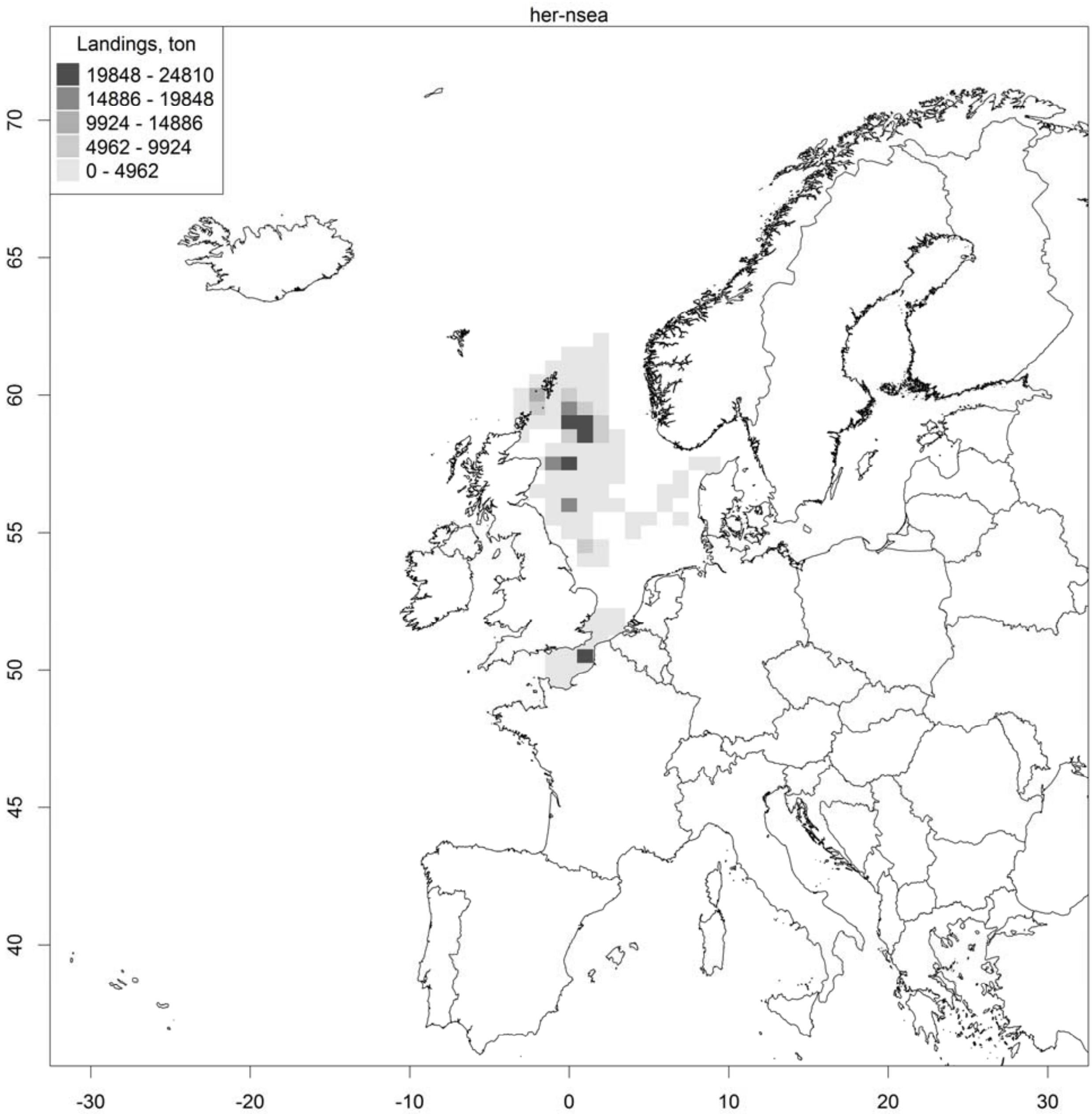


Figure 5 The North Sea herring landings by ICES square.

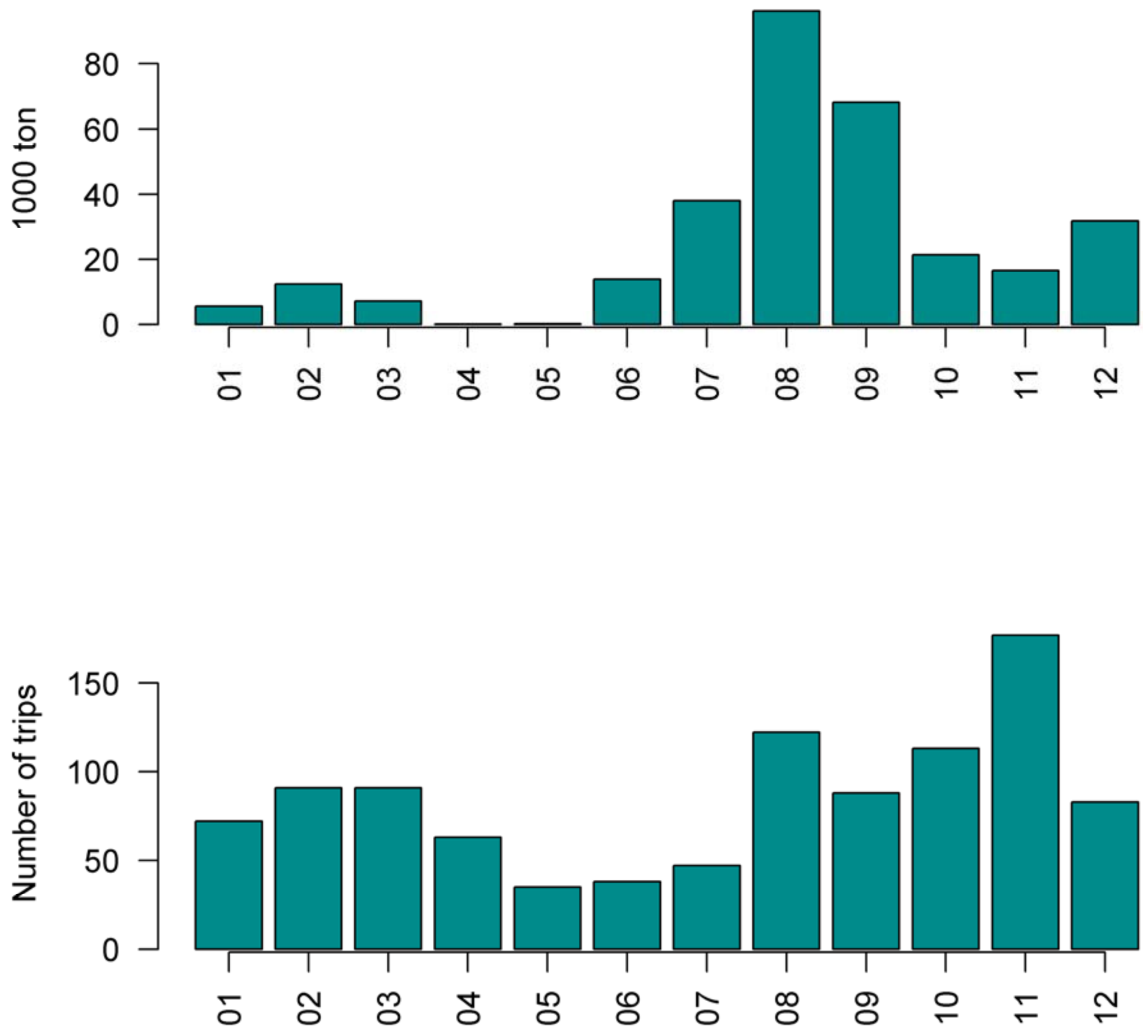


Figure 6 North Sea herring landings and number of trips by month.

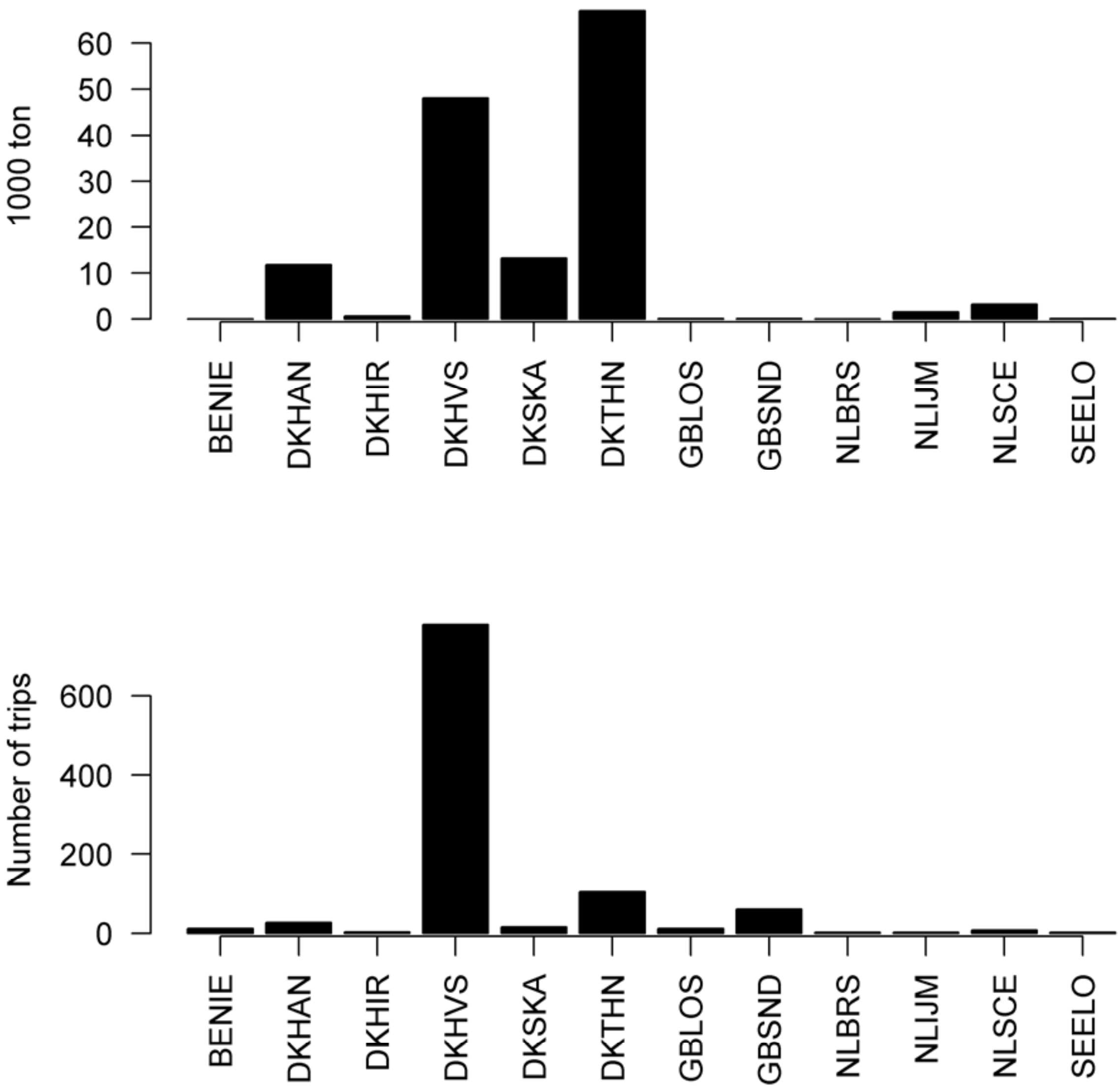
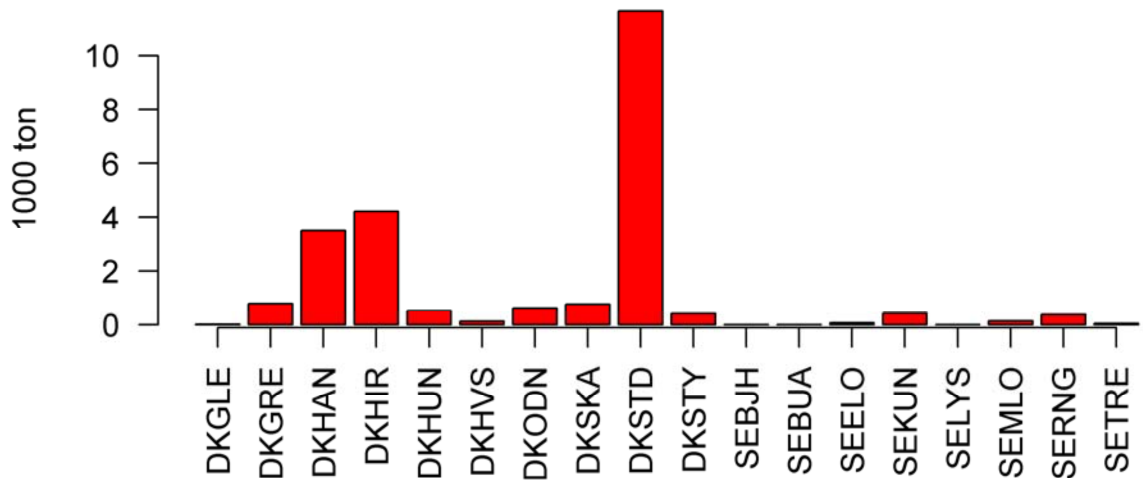


Figure 7 Tonnage landed and trips per port in 2014 for North Sea sprat. The fishery is mainly landed in Danish ports were also the number of trips landed is highest, however the port is not directly linked to the amount landed within the port and the Danish port Hvide Sande (DKHVS) has more than 700 trips in 2014 but only close to 50 000 t landed.



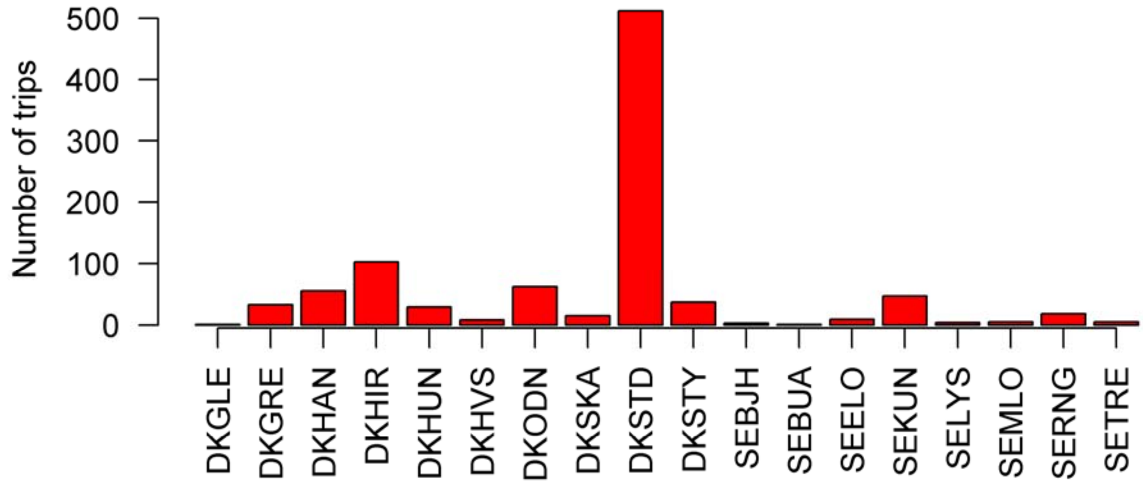


Figure 8 Tonnage landed and trips per port in 2014 for Kattegat/Skagerrak sprat. The fishery is fished with smaller vessels that in the North Sea and this can been seen by the large number of trips catching much lesser compares to the sprat stock in the North Sea. The ports with the main part of the landings in tons are very similar to the ports with the main part of the trips.

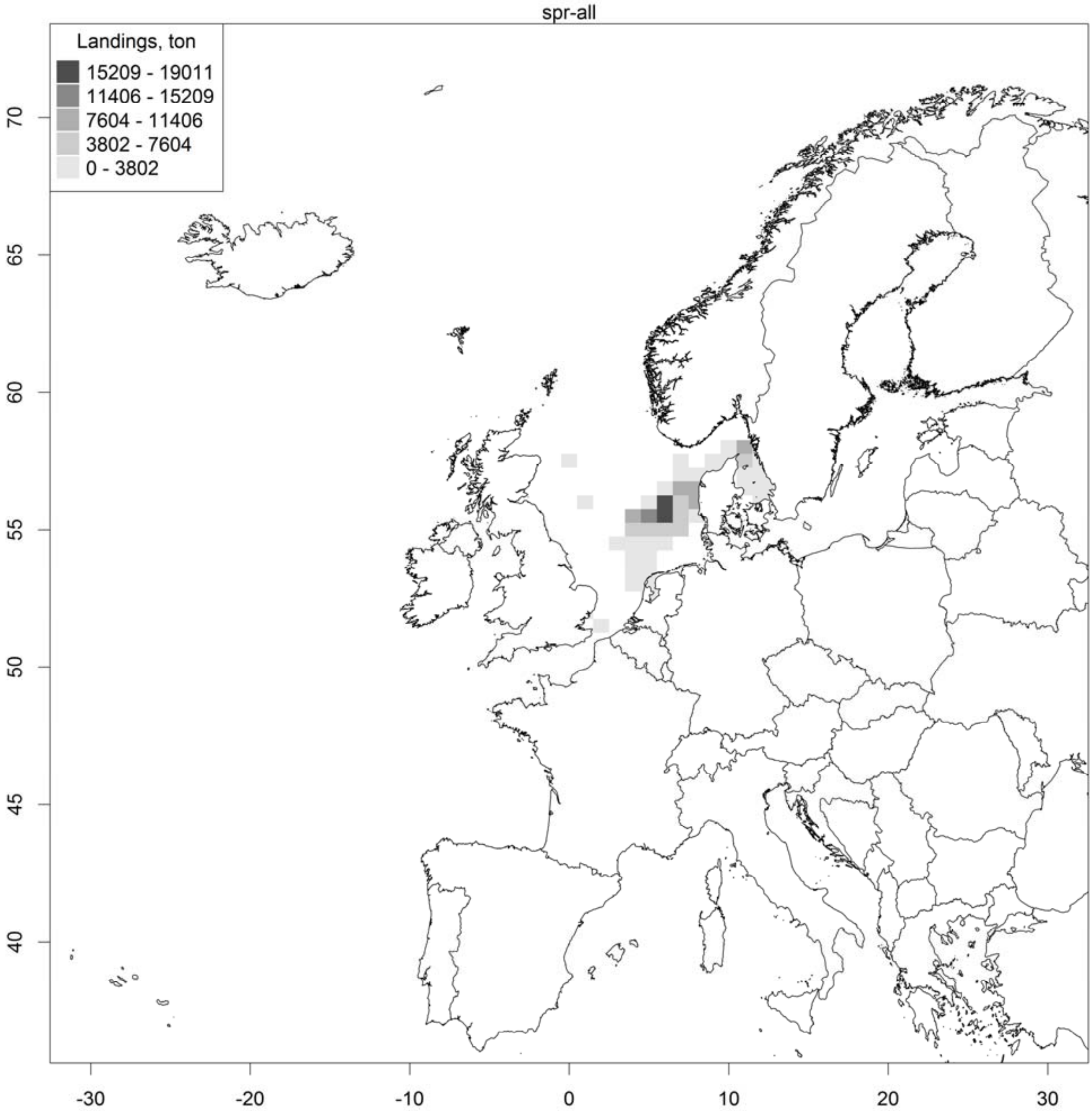


Figure 9 North Sea and Kattegat / Skagerrak landings by ICES square.

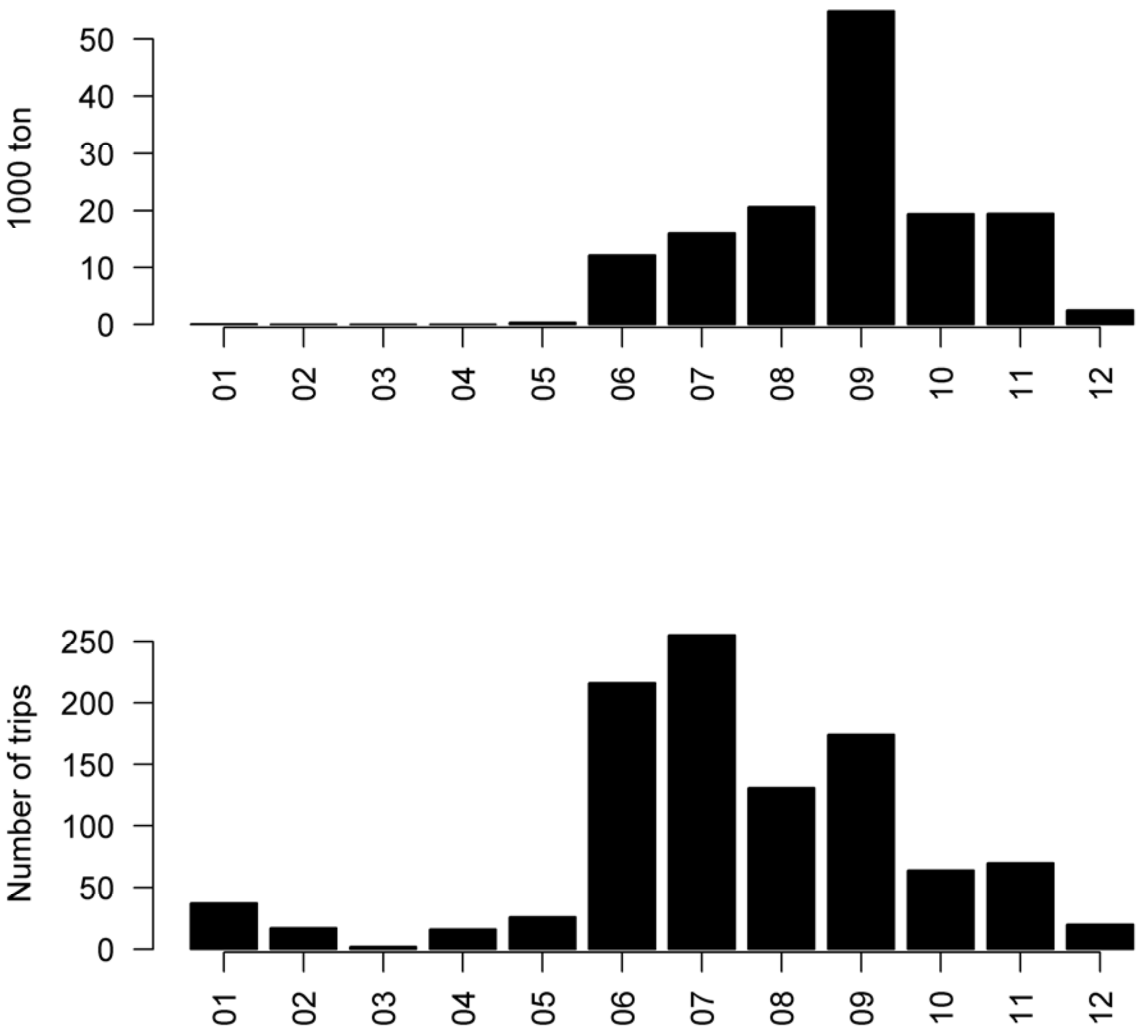


Figure 10 North Sea sprat landings and number of trips by month.

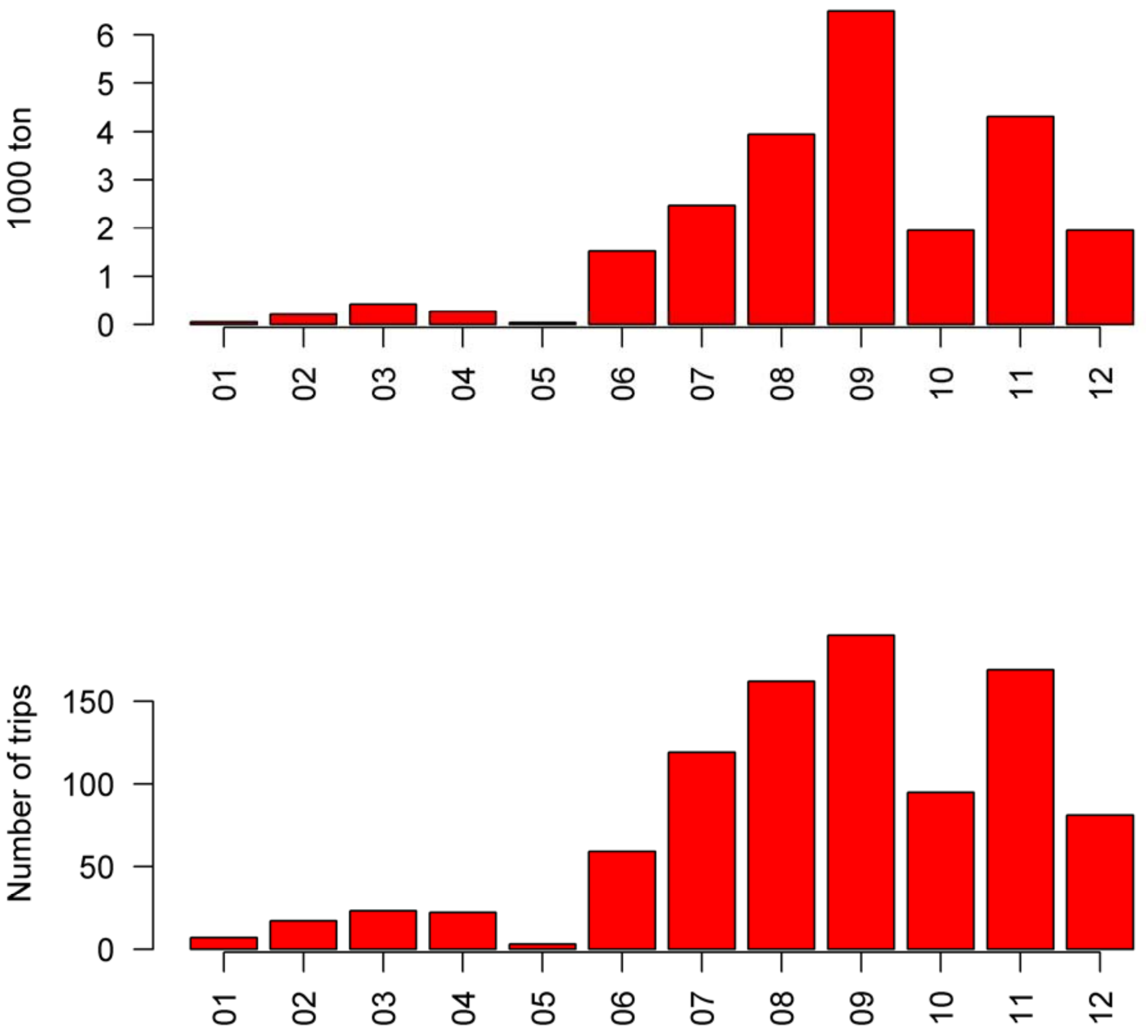


Figure 11 Kattegat/ Skagerrak sprat landings and number of trips by month.

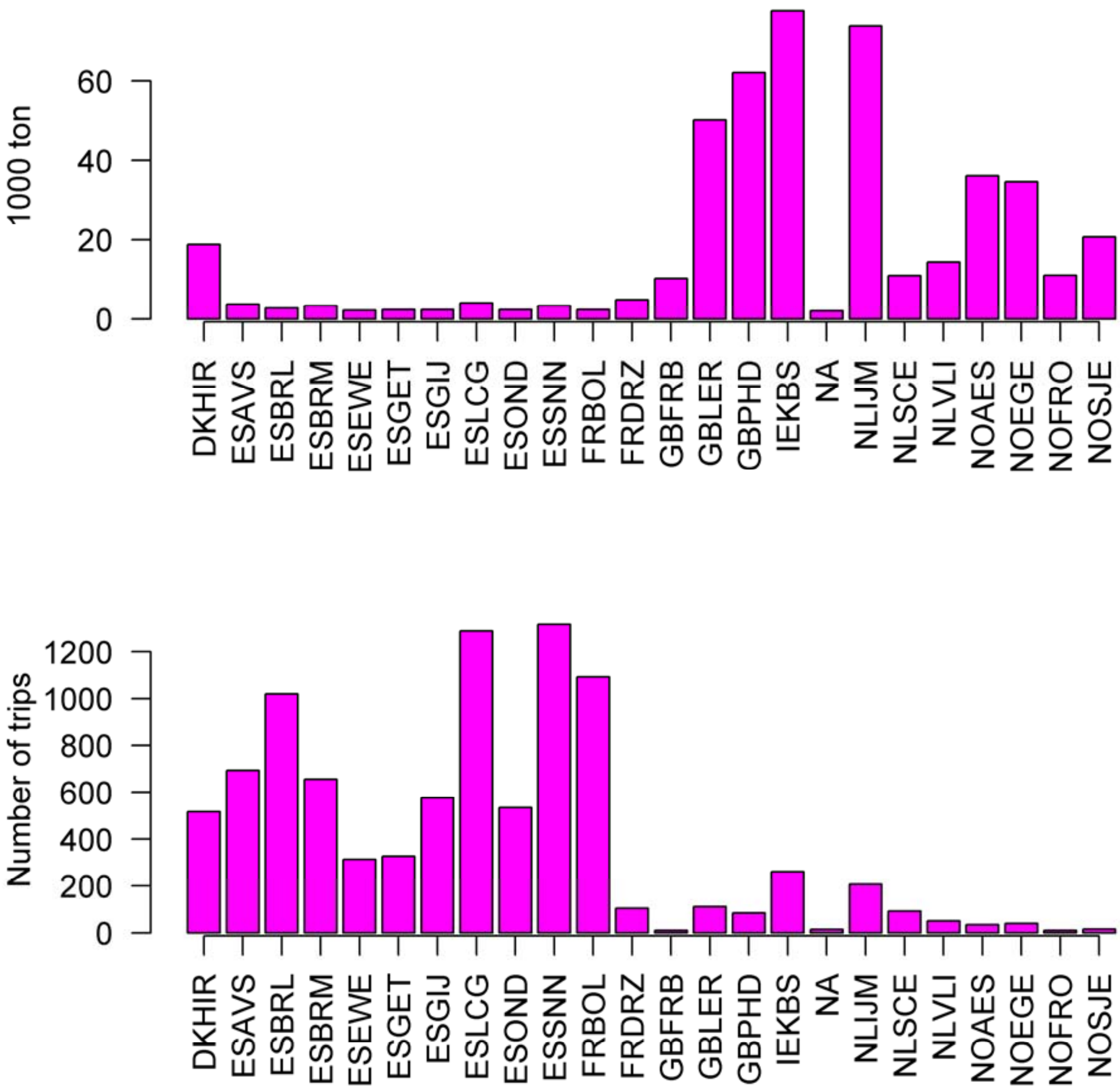


Figure 12 Tonnage landed and trips per port in 2014 for mackerel. There were 375 ports with landings of the stock, but the figures only display ports with at least 10 trips and 2 t of landings in total**.**

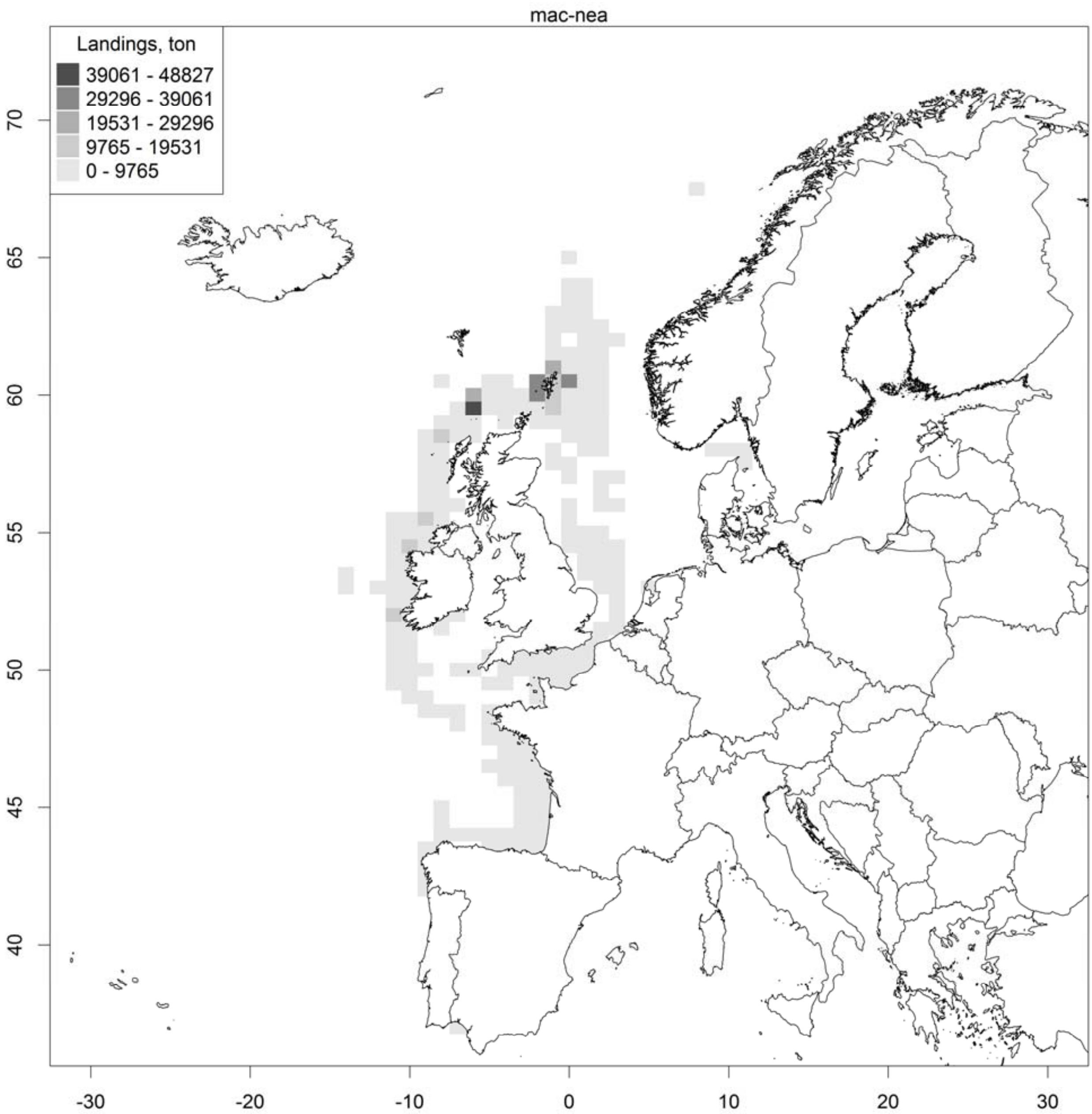


Figure 13 Mackerel landings by ICES square.

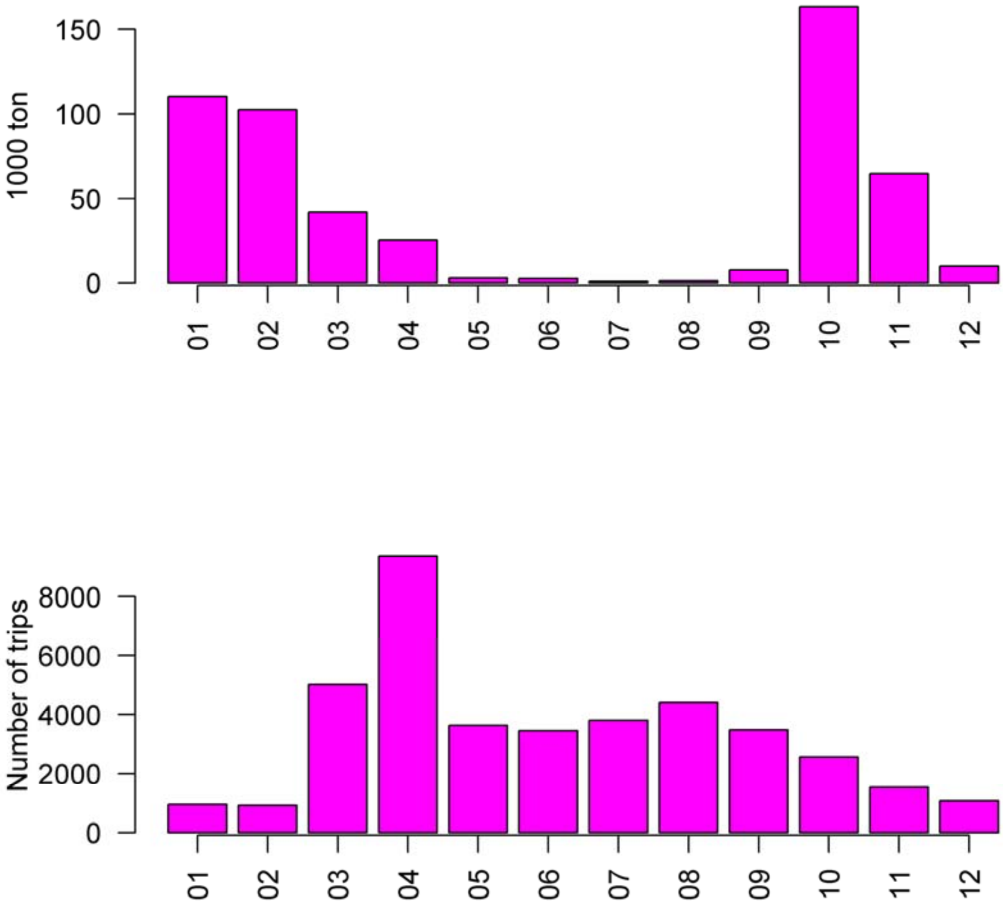


Figure 14 Mackerel landings and number of trips by month in 2014.

Table 4 The annual number of sampled market events and trips per country in 2014.

Country	Original Sampling events	Adjusted sampling events
BE	0	0
DE	3	3
DK	235	235
ES	412	6
FO	0	0
FR	60	60
GB	173	173
IE	56	56
NL	46	46
NO	0	0
PT	165	1
SE	95	95
total	1245	675

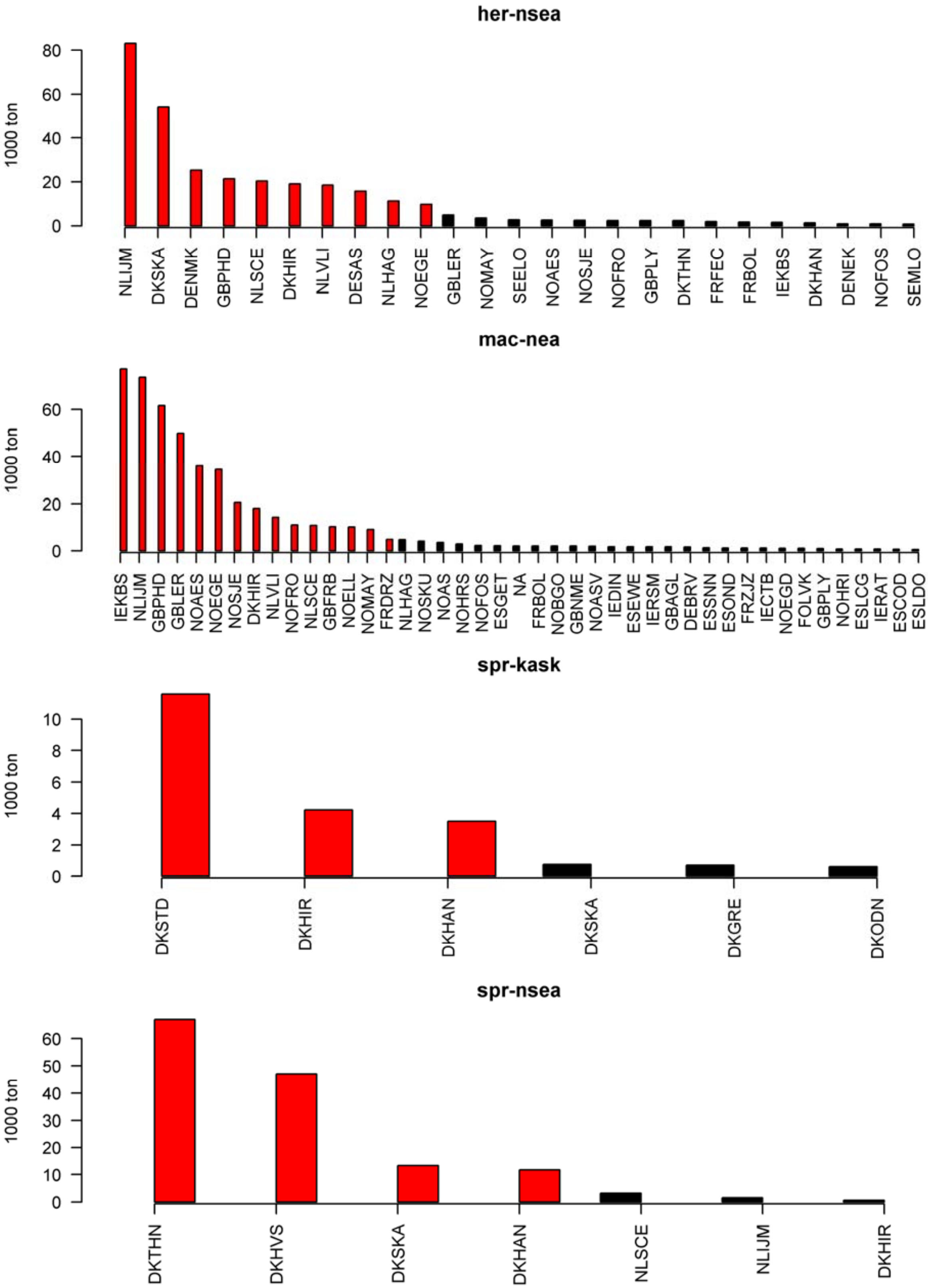


Figure 15 Grouping of ports into large and small ports. Ports were 90% of the landings per stock can be sampled (red bars). Note sampling locations with less than 500 tons are not displayed in the graph for the sake of an overview.

Table 5 Description of scenarios and results from the self-sampling designs (1A-F). Note - due to rounding of PSU's, then number of PSU's differs slightly between scenarios.

Scenario	Design	Number of sampled PSU's	Number of sampled SSU's	Number of simulations	True ton	Estimated ton	SD	SE	RSE	Design effect	Effective sample size
1A	Simple random sampling of trips – No strata. Present sampling effort used.	675	-	500	958986	959471	133192	5957	0.006	1	675
1A2	Simple random sampling of trips – No strata. Sampling effort same as scenario 1F1.	344	-	500	958986	955951	176774	7906	0.008	1	344
1A3	Simple random sampling of trips – No strata. Sampling effort same as scenario 1F2.	1316	-	500	958986	960126	83343	3727	0.004	1	1316
1A4	Simple random sampling of trips – No strata. Sampling effort same as scenario 1F3.	1113	-	500	958986	963865	95417	4267	0.004	1	1113
1B	Simple random sampling of trips – Stratified by country. Present sampling effort within country.	675	-	500	958986	959241	59154	2645	0.003	0.197	3426
1C	Simple random sampling of trips – Stratified by country. Present sampling effort weighted according to landings in strata.	671	-	500	958986	963474	66889	2991	0.003	0.252	2663
1D	Simple random sampling of trips – Stratified by vessel length group. Present sampling effort weighted according to landings in strata.	675	-	500	958986	958640	42056	1881	0.002	0.1	6750
1E	Simple random sampling of trips – Stratified by vessel length groups and country. Present within country, but weighted according to landings between vessel lengths groups.	685	-	500	958986	958545	22587	1010	0.001	0.029	23621
1F	Simple random sampling of trips – Stratified by vessel length group and country. Present sampling effort weighted according to landings in strata.	682	-	500	958986	958546	18239	816	0.001	0.019	35895
1f1	Simple random sampling of trips – Stratified by vessel length group and country. Half the present sampling effort weighted according to landings in strata.	344	-	500	958986	957642	31103	1391	0.001	0.031	11097
1f2	Simple random sampling of trips – Stratified by vessel length group and country. Twice the present sampling effort weighted according to landings in strata.	1316	-	500	958986	959182	9837	440	0.000	0.014	94000
1f3	Simple random sampling of trips – Stratified by vessel length group and country. Census for vessels above 40 m - and effort as in scenario 1f for vessel below 40 m	1113	-	500	958986	960509	15999	715	0.001	0.028	39750

Table 6 Description of scenarios and results from the port sampling designs (2A-G). * The estimated ton is biased, since landings outside EU are not sampled. Note - due to rounding of PSU's, then number of PSU's differs slightly between scenarios.

Scenario	**Design**	**Number of sampled PSU's**	**Number of sampled SSU's**	**Number of simulations**	**True ton**	**Estimated ton**	**SD**	**SE**	**RSE**	**Design effect**	**Effective sample size**
1A	Simple random sampling of trips – No strata. Present sampling effort used.	675	-	500	958986	959471	133192	5957	0.006	1	675
1A1	Simple random sampling of trips – No strata. Sampling effort same as scenario 2G.	1947	-	500	958986	956468	70450	3151	0.003	1	1947
2A	Simple random sampling of trip within port*day – No strata. Present sampling effort used.	675	675	500	958986	961408	97934	4380	0.005	0.541	1248
2B	Simple random sampling of trip within port*day – stratified by country. Present sampling effort used.	672	672	500	958986	796183*	40100	1793	0.002	0.091	7385
2c	Simple random sampling of trip within port*day – stratified by country. Present sampling effort weighted according to landings in strata.	667	667	500	958986	956841	39594	1771	0.002	0.088	7580
2d	Simple random sampling of trip within port*day – stratified by port group. Present sampling effort weighted according to landings in strata.	675	675	500	958986	957013	54914	2456	0.003	0.17	3971
2e	Simple random sampling of trip within port*day – stratified by port group and country. Present within country, but weighted according to landings between vessel lengths groups.	673	673	500	958986	807131[a]	31924	1428	0.002	0.057	11807
2f	Simple random sampling of trip within port*day – stratified by port group and country. Present sampling effort weighted according to landings in strata.	671	671	500	958986	954029	35299	1579	0.002	0.07	9586
2g	Simple random sampling of trip within port*day – stratified by port group and country. Large ports always sampled (census) small ports are random sampled four times a year. Within a marked day all large vessels are sampled (census) and a single random small vessel are sampled.	1651	1947	500	958986	959241	6409	287	0	0.008	243375

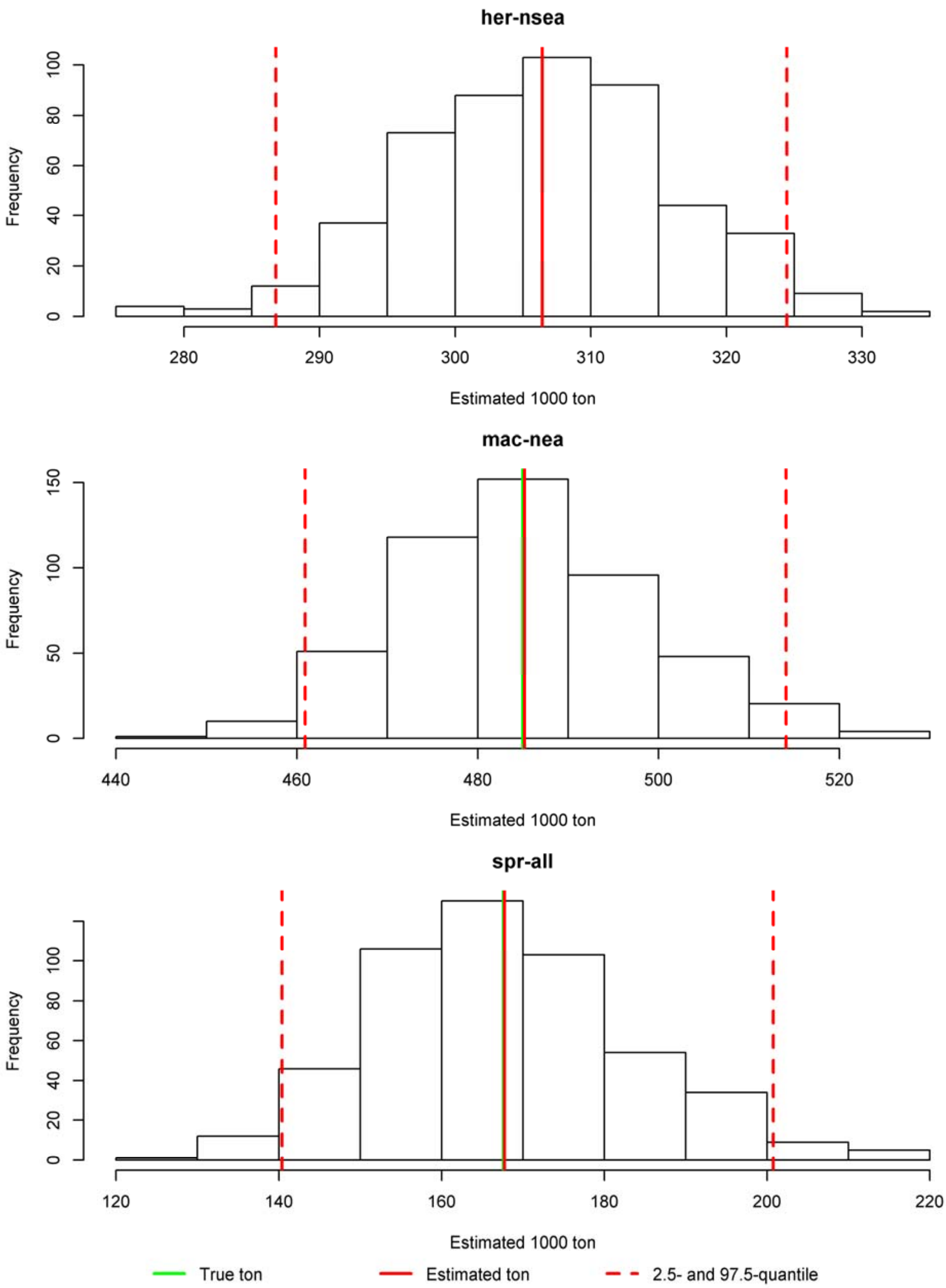
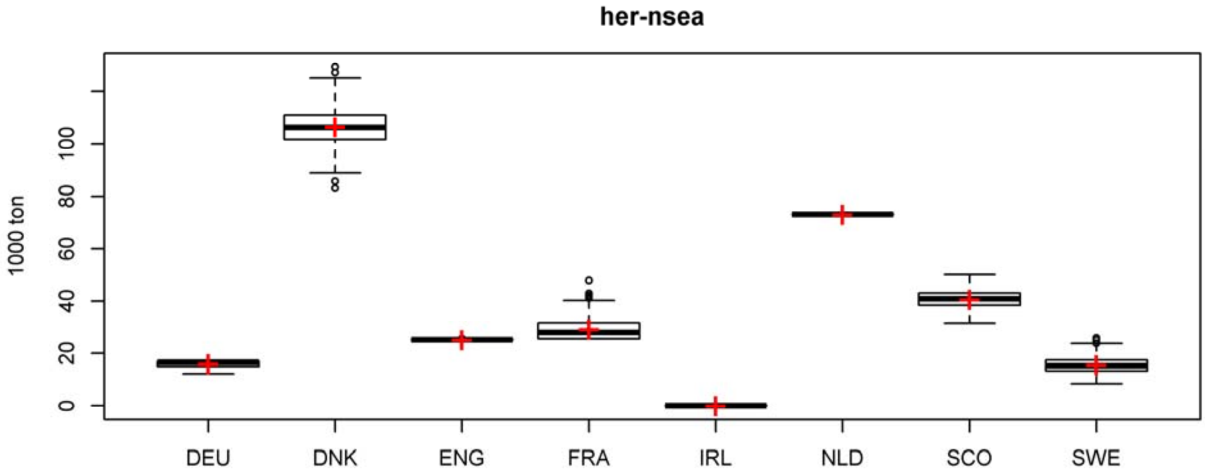
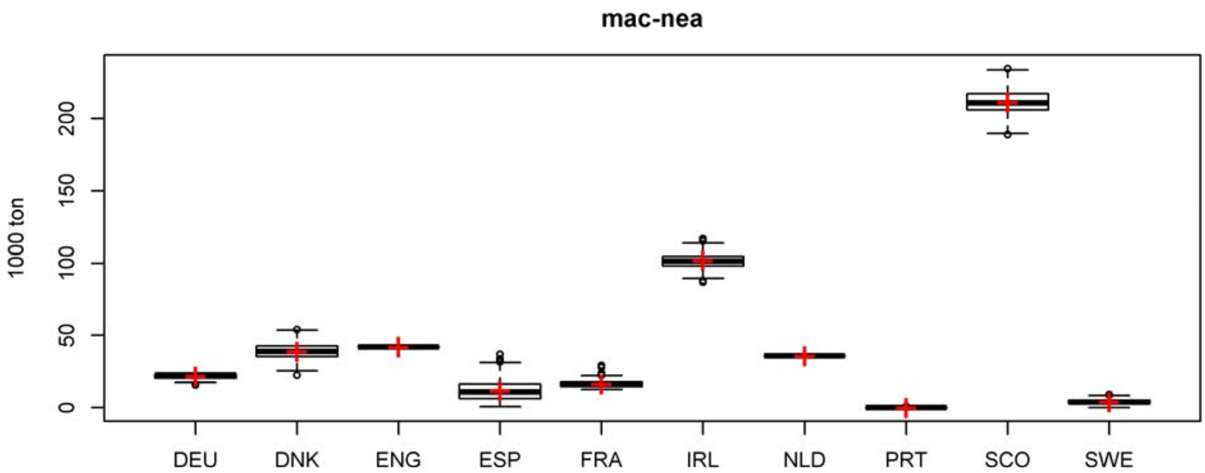


Figure 16 Result of scenario 1F with the estimated (red line) and true ton (green line) and the uncertainty around the estimate, for the three species used in this case study.





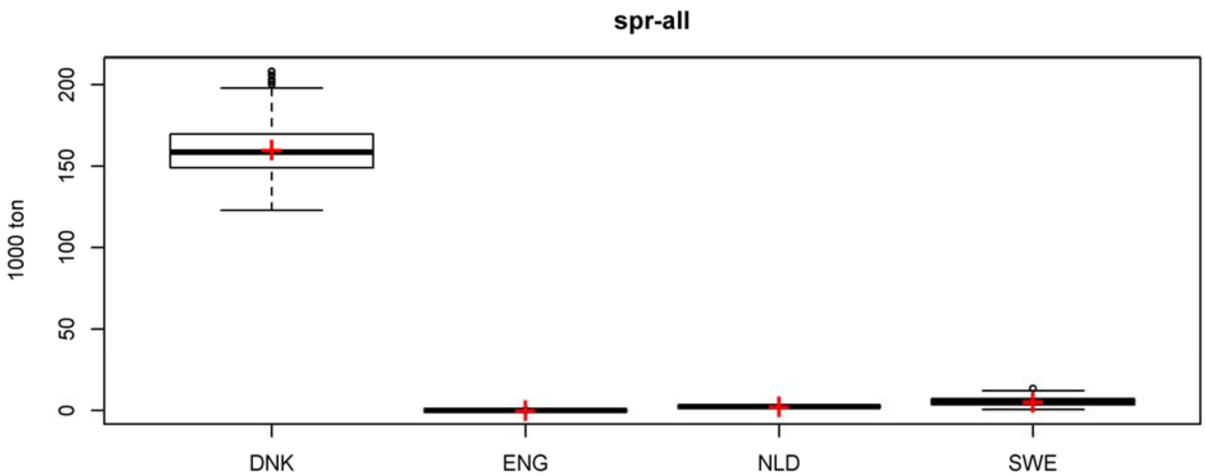


Figure 17 Coverage by country, with the true estimate (red cross), estimate median (bold black line) and uncertainty around the estimate for scenario 1F. Points are outliers.

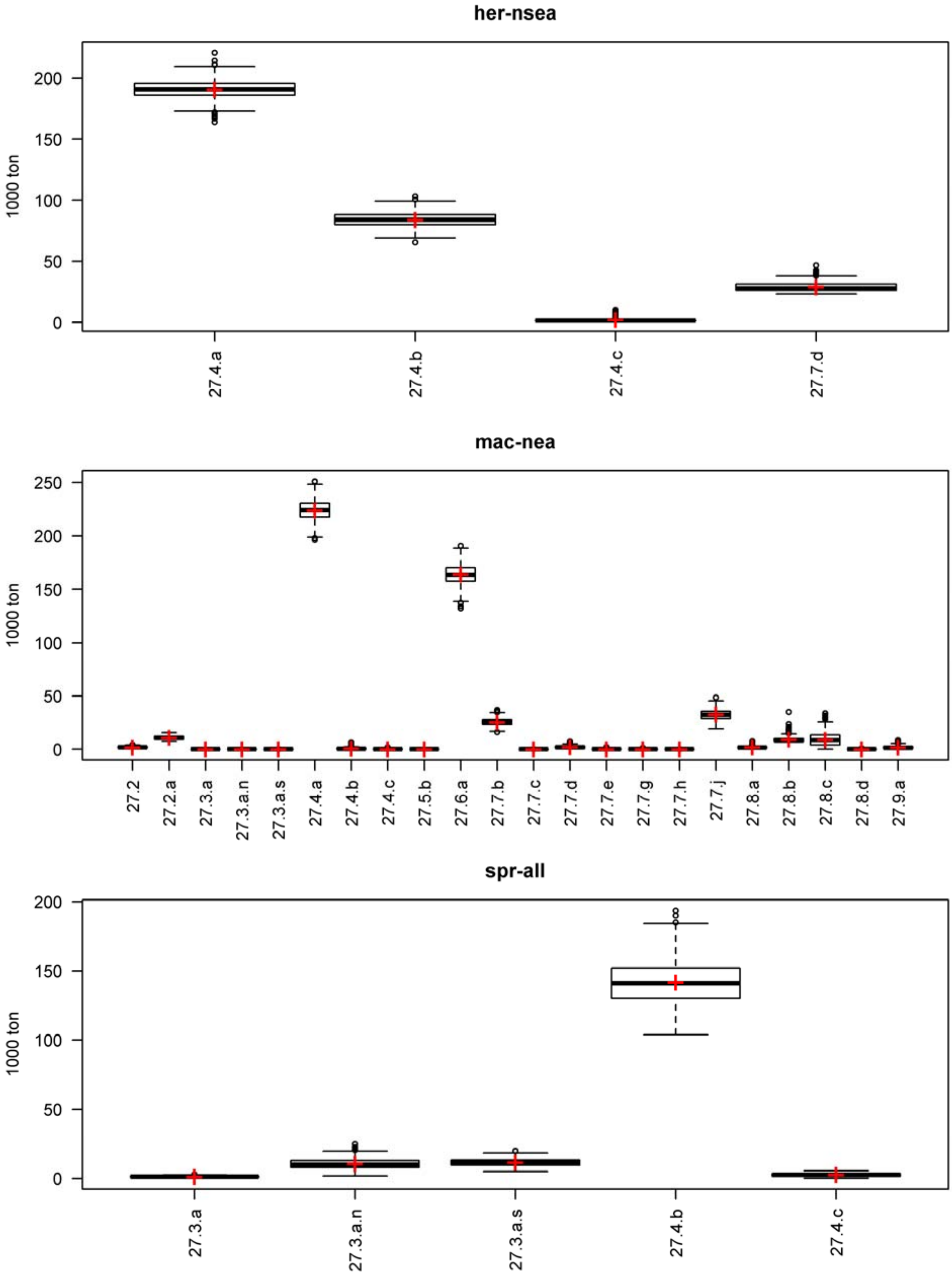


Figure 18 Coverage by area for the 3 species in the case study. The true estimate is indicated by the red cross, estimate median (bold black line) and the uncertainty around the estimate is shown for scenario 1F. Points are outliers.

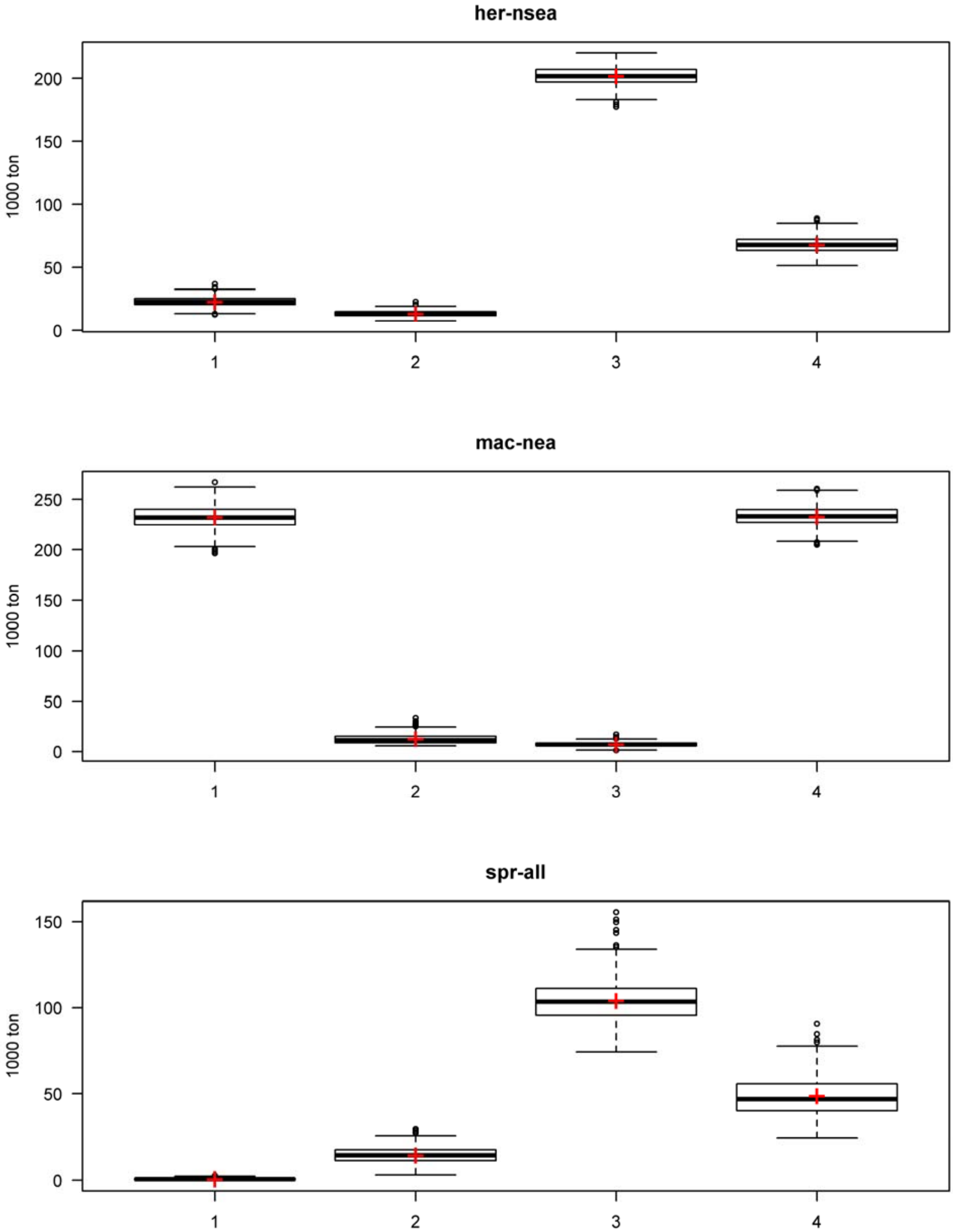


Figure 19 Coverage by quarter for the tree species in the case study, including true estimate (red cross), estimate median (bold black line) and uncertainty around the estimate for scenario 1F. Points are outliers

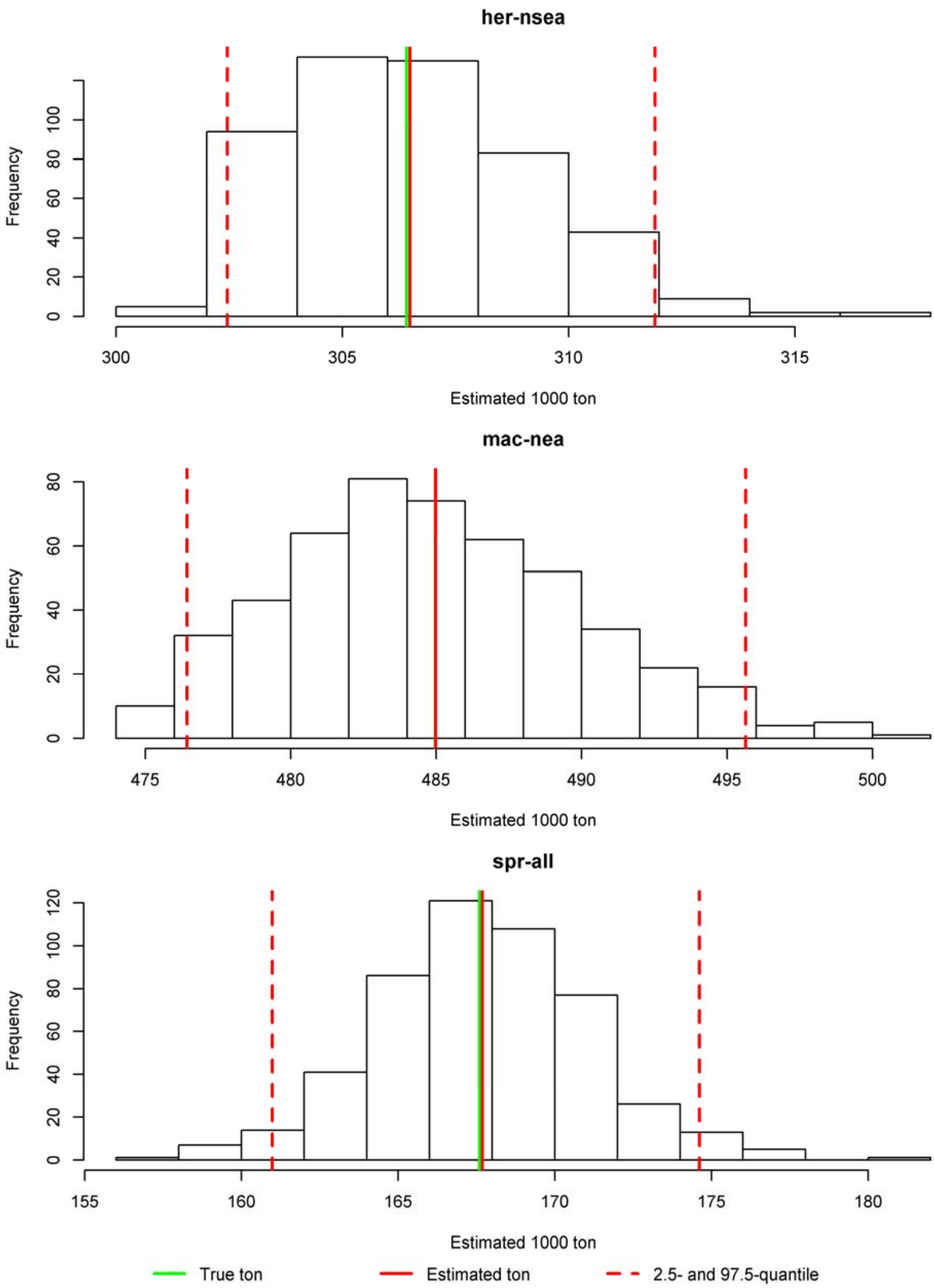


Figure 20 Result of scenario 2G with the estimated (red line) and true ton (green line) and the uncertainty around the estimate, for the three species used in this case study.

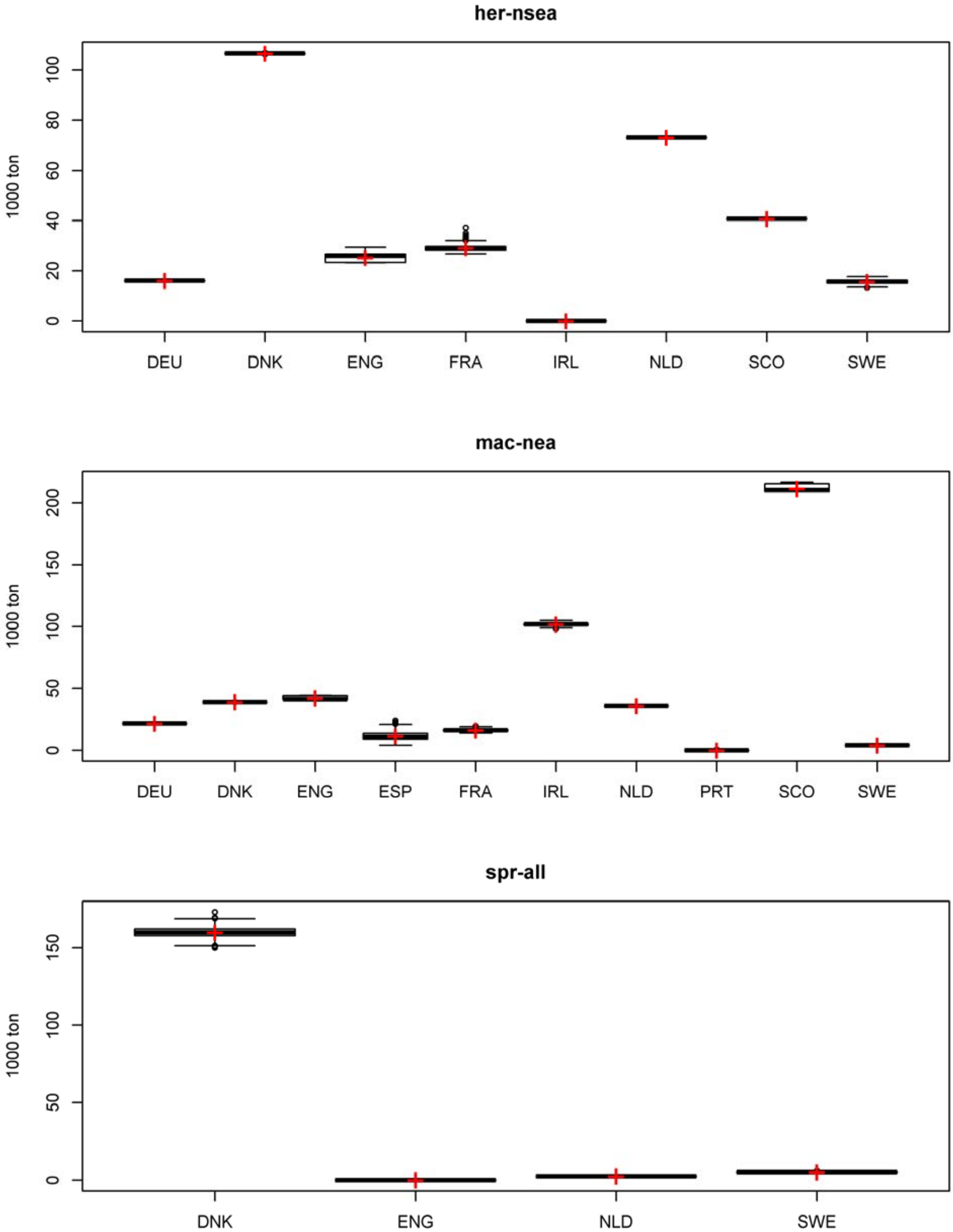


Figure 21 Coverage by country, true estimate (red cross), estimate median (bold black line) and uncertainty around the estimate for scenario 2G. Points are outliers

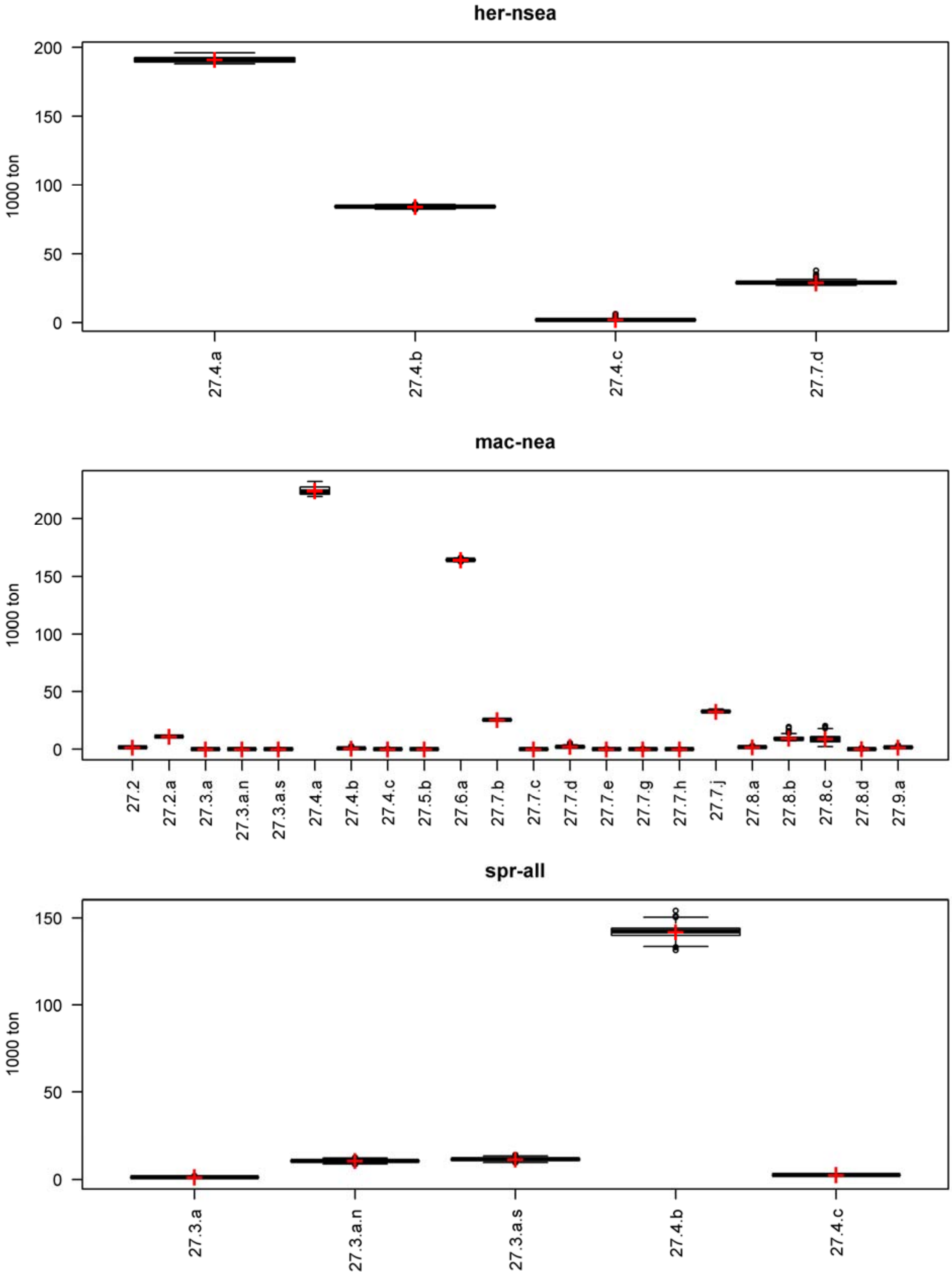


Figure 22 Coverage by area for the 3 species in the case study. The true estimate (red cross), estimate median (bold black line) and uncertainty around the estimate for scenario 2G. Points are outliers

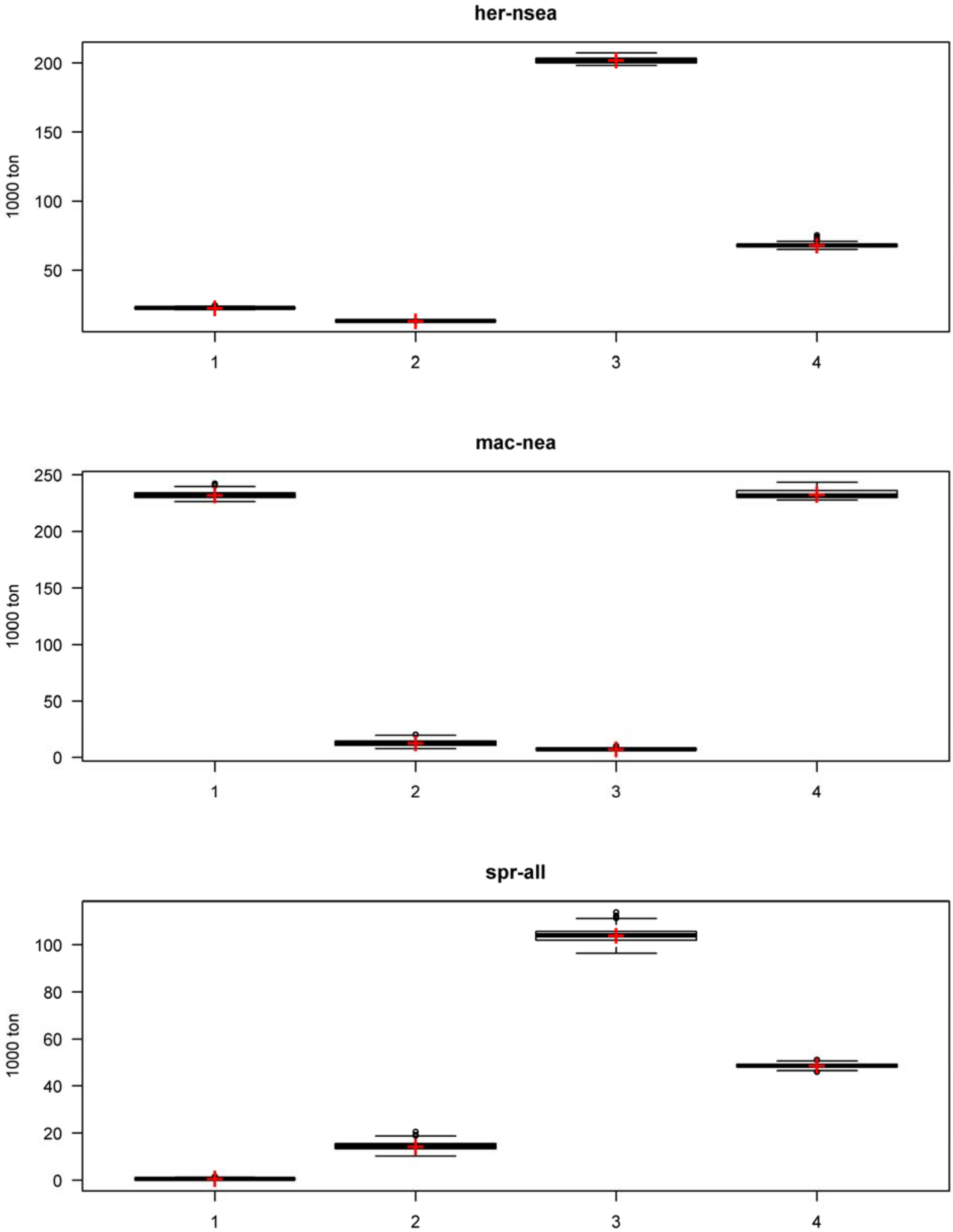


Figure 23 Coverage by quarter for the tree species in the case study, including true estimate (red cross), estimate median (bold black line) and uncertainty around the estimate for scenario 2G. Points are outliers

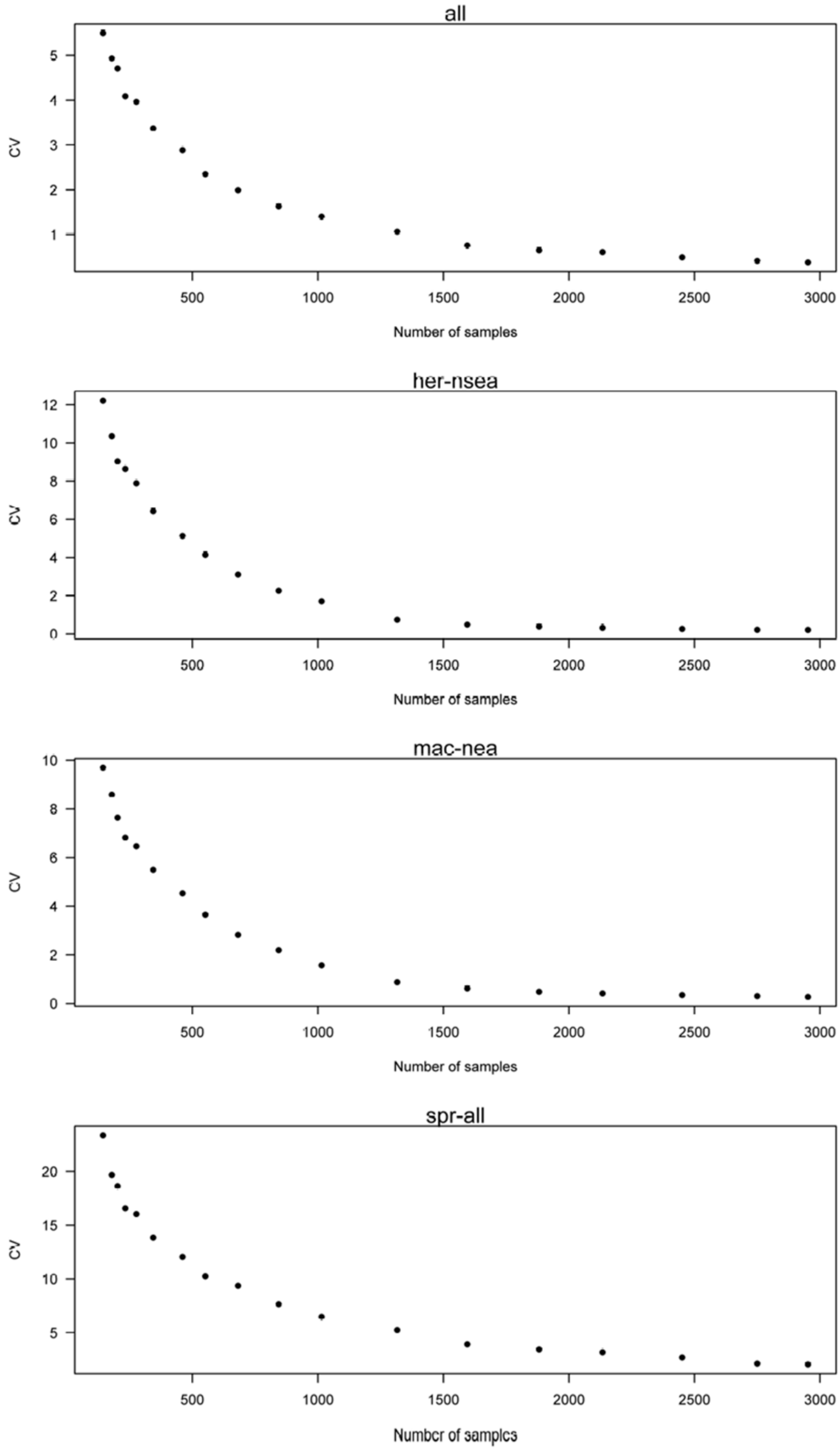


Figure 24 The figure shows the decreasing CV in correlation with the increasing in total and for the three species.

References

Borges, L., O.A. van Keeken, A.T.M. van Helmond, B. Couperus & M. Dickey-Collas, 2008. What do pelagic freezer-trawlers discard? ICES Journal of Marine Science, 65: 605-611.

Dickey-Collas, M., Engelhard, G. H., Rindorf, A., Raab, K., Smout, S., Aarts, G.,& Peck, M. A. (2013). Ecosystem-based management objectives for the North Sea: riding the forage fish rollercoaster. *ICES Journal of Marine Science: Journal du Conseil*, fst075.

Engelhard, G. H., Peck, M. A., Rindorf, A., Smout, S. C., van Deurs, M., Raab, K., & Dickey-Collas, M. (2014). Forage fish, their fisheries, and their predators: who drives whom?. *ICES Journal of Marine Science: Journal du Conseil*, *71*(1), 90-104.

Helmond van A.T.M. and Overzee, van H.M.J. 2010b. Can pelagic freezer trawlers reduce discarding? Working Document to the Herring assessment working Group. 4pp.

ICES. 2014. Report of the Working Group on Widely Distributed Stocks (WGWIDE). 26 August - 1 September 2014. ICES Headquarters, Copenhagen, Denmark. ICES CM 2014/ACOM:15. 945 pp.

Jansen, T., Gislason, H. 2011. Temperature affects the timing of spawning and migration of North Sea mackerel. Continental Shelf Research. 31: 64-72.

Napier, I.R., Newton, A.W. and Toreson, R. 1999. Investigation of the Extent and Nature of Discarding from Herring and Mackerel Fisheries in ICES Sub-Areas IVa and VIa. Final Report. EU Study Contract Report 96/082. North Atlantic Fisheries College, Shetland Islands, UK. June 1999.

Napier, I.R., Robb, A. and Holst, J. 2002. Investigation of Pelagic Discarding. Final Report. EU Study Contract Report 99/071. North Atlantic Fisheries College and the FRS Marine Laboratory. August 2002.

Appendix 1 Data scrutinising

After data submission by each institute, data were checked and standardised across institutes and collated in one large dataset to be used in simulations. For case study 1 – the pelagic fisheries, the following fields were corrected, added or filtered to provide a final target population for the simulations.

Corrected

The following variables in the dataset was carefully checked and corrected if wrong.

- On-shore sample location/ arrival location/ departure location (onShoreSampLoc/ arvLoc/ depLoc): Wrong codes were in general corrected. There are still NA's in the data due to difficulties in finding the correct codes. NA's for locations with very small amount of landings were in general accepted.
- On-shore sample location (onShoreSampLoc): If not filled in then filled with arrival location (arvLoc).
- Métier (foCatEu6): There were a lot of wrong métiers in the data. Gear and target species were checked, but wrong mesh size ranges were not corrected, since mesh size range is not used in the present study.
- Vessel length class (vslLenCls): Wrong codes were corrected.
- Vessel flag country (vslFlgCtry): Renamed all English country codes to ENG.

Added

The following variables were added to the dataset.

- ton: landWt/1000
- year: From arrival date (arvDate).
- quarter: From arrival date.
- month: From arrival date.
- week: From arrival date.
- daysAtSea: Difference between departure and arrival dates.
- Target species (sppTarget): Most fished species on trip.
- Target stock (stockTarget): Most fished stock.
- Target species group (sppGrpfoCatEu6): Species group (SPF/Other) deduced from foCatEu6.
- Market day (marketDay): On-shore sample location combined with arrival date (onShoreSampLoc||arvDate).
- Working group area (wgArea): Areas used in assessment
    - Sprat - IVa, IVbW, IVbE, IVc (HAWG areas) - the model runs on rectangles.
    - Herring - IVaW, IVaE, IVb, IVc and VIId for HUC and IVaW, IVaE, IVbW, IVbEW, IVbEE, IVc and VIId
    - for IND (HAWG areas)
    - Mackerel - special areas?
- On-shore sampling country (onShoreSampCtry): Country deduced from On-shore sample location.

Filtered (finding the target population)

To conduct a targeted sampling scheme the main challenges is to track the landings going to the factories, since there is no clear distinction in the data used for this project – and maybe not even in data at the national level.

Criteria used for filtering

1. Only data from 2014 have been used.
2. Trips where none of the 4 stocks (focused on in the CS) have been caught were deleted.
3. Areas outside 27 have been deleted.
4. Only trips where the target stock was one of the 4 in question have been selected. The trips landing to factories are presumed to target a certain stock/species.
5. Trips not targeting small pelagic have been deleted and only trips fishing with trawls or purse seiners were kept. This step excludes mixed fisheries and passive gears assuming the catch from these fisheries are going to the auctions. In the original data there were 185 different métiers, which had caught at least one of the 4 stocks, see figure a.

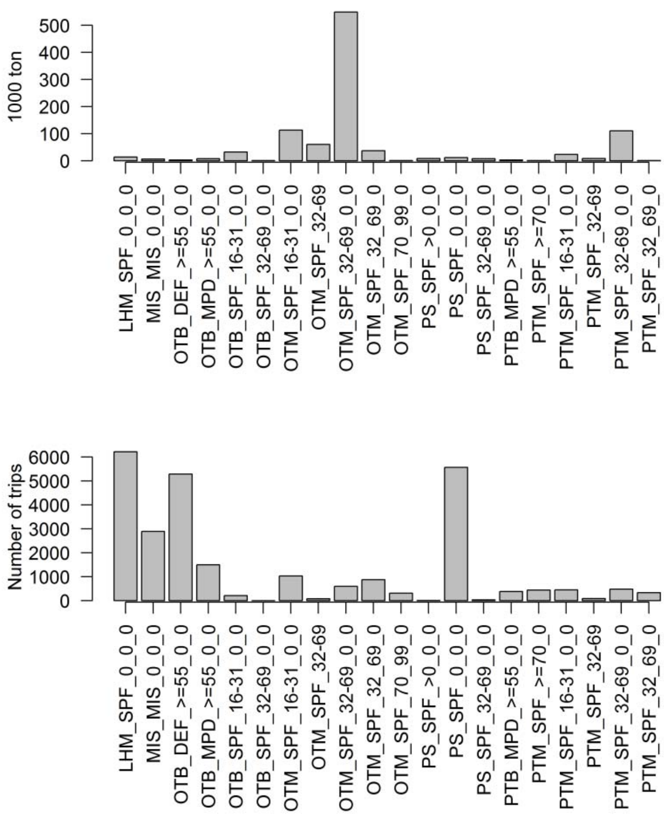


Figure a Amount of landing in tons by métier and numbers of trips conducted by métier for the 4 stocks analysed in this case study in the 2014. Métiers with less than 5 trips and 1 ton of landings are not displayed in this graph to keep an overview.

Consequences of filtering

Although the dataset was truncated and thereby many trips were sorted out of the final dataset only a very small amount of the landed tons were filtered out (Tables a, b, and c).

Table a Difference in number of trips and tons after deleting trips not targeting small pelagic and only keeping trips fishing with trawls or purse seiners.

year	stock	Trips before*	Ton before	Deleted trips	Deleted ton	Deleted trips (%)	Deleted ton (%)
2014	her-nsea	1,044	311,658	286	2,570	27	1
2014	mac-nea	40,483	533,991	30,912	34,770	76	7
2014	spr-kask	948	23,644	15	193	2	1
2014	spr-nsea	1,028	145,600	43	1,171	4	1

*Note trips can be counted more than ones if more than one stock has been caught on a trip.

Table b Difference in number of trips and tons after deleting trips not targeting one of the stocks.

year	stock	Trips before	Ton before	Deleted trips	Deleted ton	Deleted trips (%)	Deleted ton (%)
2014	her-nsea	758	309,088	42	2,239	6	1
2014	mac-nea	9,574	499,221	2,038	12,195	21	2
2014	spr-kask	933	23,450	26	282	3	1
2014	spr-nsea	985	144,429	NA	NA	NA	NA

Table c Difference in number of trips and tons after deleting trips with no information of sampling location.

year	stock	Trips before	Ton before	Deleted trips	Deleted ton	Deleted trips (%)	Deleted ton (%)
2014	her-nsea	716	306,849	2	430	6	1
2014	mac-nea	7,536	487,025	2	2056	27	3
2014	spr-kask	907	23,168	NA	NA	NA	NA
2014	spr-nsea	985	144,429	NA	NA	NA	NA

Appendix 2 Number of vessels per vessel length group

Table d Numbers of vessels and trips by country and vessel length group.

Country and vessel Length group	Number of vessels	Number of trips
Germany <40 m	1	2
Germany >40 m	5	32
Denmark <40 m	54	1566
Denmark >40 m	31	321
England <40 m	4	147
England >40 m	4	48
Spain <40 m	365	4784
Spain >40 m	1	1
France <40 m	131	1541
France >40 m	2	27
Ireland <40 m	36	216
Ireland >40 m	21	173
Netherlands <40 m	0	0
Netherlands >40 m	9	39
Portugal <40 m	92	531
Portugal >40 m	0	0
Scotland <40 m	1	17
Scotland >40 m	23	285
Sweden <40 m	11	112
Sweden >40 m	10	71
Total	801	9913

Appendix 3 Number of ports per port group

Table e Numbers of sampling locations and market-days by country and sampling location group

Sampling country	Location group	Number of sampling locations	Number of market days
Germany	Large	2	53
Germany	Small	2	5
Denmark	Large	6	564
Denmark	Small	5	135
Spain	Small	48	1750
Faroe Island	Small	1	2
France	Large	1	16
France	Small	30	926
Great Britain	Large	3	175
Great Britain	Small	9	114
Ireland	Large	1	80
Ireland	Small	15	68
Netherlands	Large	4	98
Norway	Large	6	122
Norway	Small	10	28
Portugal	Small	13	245
Sweden	Small	7	52

Appendix 4 Sampling locations per country

Table f Detailed list of sampling locations per country and location group

Sampling country	Location group	Sample location code	Sampling location name
Germany	Large	DENMK	NEU MUKRAN
Germany	Large	DESAS	SASSNITZ
Germany	Small	DEBRV	BREMERHAVEN
Germany	Small	DENEK	Neuenkirchen, Rügen
Denmark	Large	DKHAN	Hanstholm
Denmark	Large	DKHIR	Hirtshals
Denmark	Large	DKHVS	Hvide Sande
Denmark	Large	DKSKA	Skagen
Denmark	Large	DKSTD	Strandby
Denmark	Large	DKTHN	Thyborøn
Denmark	Small	DKGLE	Gilleleje
Denmark	Small	DKGRE	Grenaa
Denmark	Small	DKHUN	Hundested
Denmark	Small	DKODN	Odden
Denmark	Small	DKSTY	Strandby, Langeland
Spain	Small	ESAVS	Aviles
Spain	Small	ESBBE	NA
Spain	Small	ESBRL	Burela
Spain	Small	ESBRM	Bermeo
Spain	Small	ESBUE	Bueu
Spain	Small	ESCAD	Cádiz
Spain	Small	ESCBD	Cambados
Spain	Small	ESCIO	Cillero
Spain	Small	ESCNO	Cariño
Spain	Small	ESCOD	Colindres
Spain	Small	ESDAA	Carreira-Aguiño
Spain	Small	ESEGR	El Grove
Spain	Small	ESEWE	Fuenterrabia
Spain	Small	ESFNE	Finisterre
Spain	Small	ESGET	NA
Spain	Small	ESGIJ	Gijón
Spain	Small	ESHUV	Huelva
Spain	Small	ESIAS	Camariñas
Spain	Small	ESLAX	Lage
Spain	Small	ESLCG	La Coruña
Spain	Small	ESLDO	Laredo
Spain	Small	ESLS3	Lastres
Spain	Small	ESMAI	Malpica de Bergantiños
Spain	Small	ESMEI	Meira
Spain	Small	ESMOA	Moaña
Spain	Small	ESMPG	Marin
Spain	Small	ESMRS	Muros
Spain	Small	ESNOA	Noia
Spain	Small	ESOND	Ondarroa
Spain	Small	ESPAS	Pasajes
Spain	Small	ESPFT	NA
Spain	Small	ESPRT	Portosín
Spain	Small	ESPSM	Puerto de Santa María
Spain	Small	ESPTD	Puentedeume
Spain	Small	ESRAN	NA
Spain	Small	ESSAD	Sada
Spain	Small	ESSBA	Sanlucar De Barrameda
Spain	Small	ESSDR	Santander
Spain	Small	ESSNI	Santa Eugenia de Riveira
Spain	Small	ESSNN	Santoña
Spain	Small	ESSNR	Santurce
Spain	Small	ESSVB	San Vicente Barquera

Spain	Small	ESVEM	Villajuan
Spain	Small	ESVGO	Vigo
Spain	Small	ESVIV	Vivero
Spain	Small	ESZGA	Isla Cristina
Spain	Small	ESZJN	Portonovo
Spain	Small	ESZJY	Punta Umbria
Faroe Islands	Small	FOLVK	Leirvik
France	Large	FRDRZ	Douarnenez
France	Small	FR2GX	Barfleur
France	Small	FRARC	Arcachon
France	Small	FRBAY	Bayonne
France	Small	FRBOL	Boulogne-sur-Mer
France	Small	FRCER	Cherbourg
France	Small	FRCOC	Concarneau
France	Small	FRDPE	Dieppe
France	Small	FRDSM	Dives-sur-Mer
France	Small	FRELS	Étaples
France	Small	FRFEC	Fécamp
France	Small	FRGFR	Granville
France	Small	FRGVC	Guilvinec
France	Small	FRHEN	Hendaye
France	Small	FRHON	Honfleur
France	Small	FRLC5	La Cotinière
France	Small	FRLEH	Le Havre
France	Small	FRLOC	Loctudy
France	Small	FRLRH	La Rochelle
France	Small	FRLSO	Les Sables-d'Olonne
France	Small	FRLTR	Le Tréport
France	Small	FROUI	Ouistreham
France	Small	FRQUI	Quiberon
France	Small	FRQUY	Erquy
France	Small	FRRTB	Port-en-Bessin-Huppain
France	Small	FRTBE	La Turballe
France	Small	FRTJE	Trouville-sur-Mer
France	Small	FRVHH	NA
France	Small	FRXGV	Saint-Gilles-Croix-de-Vie
France	Small	FRXSG	NA
France	Small	FRZJZ	Saint-Jean-de-Luz
Great Britain	Large	GBFRB	Fraserburgh
Great Britain	Large	GBLER	Lerwick
Great Britain	Large	GBPHD	Peterhead
Great Britain	Small	GBAGL	Ardglass
Great Britain	Small	GBCRR	Cromer
Great Britain	Small	GBLDY	Londonderry
Great Britain	Small	GBLOS	Leigh-on-Sea
Great Britain	Small	GBNME	Northmavine
Great Britain	Small	GBPLY	Plymouth
Great Britain	Small	GBSND	Southend
Great Britain	Small	GBULL	Ullapool
Great Britain	Small	GBWPT	Warrenpoint
Ireland	Large	IEKBS	KILLYBEGS
Ireland	Small	IEARR	ARRANMORE
Ireland	Small	IEBCV	BALLYCROVANE
Ireland	Small	IEBLN	BALLYCOTTON
Ireland	Small	IEBTM	BALTIMORE
Ireland	Small	IECTB	Castletown Bearhaven
Ireland	Small	IEDIN	Dingle
Ireland	Small	IEDNM	DUNMORE EAST
Ireland	Small	IEGRE	Greencastle
Ireland	Small	IEHOW	HOWTH
Ireland	Small	IEKLA	KILLALA
Ireland	Small	IEPOY	PORTURLIN/PORTACLOY
Ireland	Small	IERAT	RATHMULLAN

Ireland	Small	IERSM	ROS A MHIL
Ireland	Small	IETEE	TEELIN
Ireland	Small	IEUNI	UNION HALL
Netherlands	Large	NLHAG	's-Gravenhage (Den Haag)
Netherlands	Large	NLIJM	IJmuiden/Velsen
Netherlands	Large	NLSCE	Scheveningen
Netherlands	Large	NLVLI	Vlissingen
Norway	Large	NOAES	Ålesund
Norway	Large	NOEGE	EGERSUND
Norway	Large	NOELL	Ellingsøy
Norway	Large	NOFRO	Florø
Norway	Large	NOMAY	MALOEY
Norway	Large	NOSJE	Selje
Norway	Small	NO999	NA
Norway	Small	NOAS	NA
Norway	Small	NOASV	AUSTEVOLL
Norway	Small	NOBGO	BERGEN
Norway	Small	NOEGD	Eigersund
Norway	Small	NOFOS	Fosnavåg
Norway	Small	NOHRI	Hareid
Norway	Small	NOHRS	NA
Norway	Small	NOSKU	Skudeneshavn
Norway	Small	NOSMO	Smøla
Portugal	Small	PTAVE	Aveiro
Portugal	Small	PTFDF	Figueira da Foz
Portugal	Small	PTLOS	Lagos
Portugal	Small	PTMAT	Matosinhos
Portugal	Small	PTOLH	Olhão
Portugal	Small	PTPDV	Povoa de Varzim
Portugal	Small	PTPEN	Peniche
Portugal	Small	PTPRM	Portimao
Portugal	Small	PTQRT	NA
Portugal	Small	PTSAG	Sagres
Portugal	Small	PTSET	Setúbal
Portugal	Small	PTSIE	Sines
Portugal	Small	PTSSB	Sesimbra
Sweden	Small	SEELO	Ellös
Sweden	Small	SEFJA	Fjällbacka
Sweden	Small	SEKUN	Kungshamn
Sweden	Small	SELYS	Lysekil
Sweden	Small	SEMLO	Mollösund
Sweden	Small	SERNG	Rönnäng
Sweden	Small	SETRE	Träslövsläge

Appendix 5 Detailed results from scenarios tested

Table 1 Sampling effort for 5 different scenarios in 1A.

Scenario	Strata	Total number of trips	Numbers of planned sampling events
1A	Not available	9913	675
1A1	Not available	9913	1947 (similar to 2G)
1A2	Not available	9913	50% present effort
1A3	Not available	9913	200% present effort
1A4	Not available	9913	Census for all vessel >40 meter – same effort as in 1F for vessels <40 meter

Table 2. Strata, numbers of PSU and sample events per strata for scenario 1B.

Strata: Vessel flag country	Total number of trips	Numbers of planned sampling events
Germany	34	3
Denmark	1887	236
England	195	43
Spain	4785	7
France	1568	61
Ireland	389	57
Netherlands	39	39
Portugal	531	2
Scotland	302	132
Sweden	183	95
Total	9913	675

Table 3. Strata, numbers of PSU and sample events per strata for scenario 1C.

Strata: Vessel flag country	Total number of trips	Numbers of planned sampling events
Germany	34	31
Denmark	1887	219
England	195	52
Spain	4785	12
France	1568	36
Ireland	389	76
Netherlands	39	39
Portugal	531	4
Scotland	302	181
Sweden	183	21
Total	9913	671

Table 4. Strata, numbers of PSU and sample events per strata for scenario 1D.

Strata: Vessel flag country	Total number of trips	Numbers of planned sampling events	Numbers of archived sampling Events
Vessel length < 40m	8916	87	87
Vessel length > 40m	997	588	588

Table 5. Strata, numbers of PSU and sample events per strata for scenario 1E. Each country has 2 strata for vessels above and below 40 meters.

Strata: Vessel flag country	Total number of trips	Numbers of planned sampling events
Germany < 40 m	2	2
Germany > 40 m	32	7
Denmark < 40 m	1566	64
Denmark > 40 m	321	179
England < 40 m	147	6
England > 40 m	48	46
Spain< 40 m	4784	10

Spain> 40 m	1	0
France< 40 m	1541	14
France> 40 m	27	27
Ireland< 40 m	216	13
Ireland> 40 m	173	51
Netherlands < 40 m	0	0
Netherlands > 40 m	39	39
Portugal< 40 m	531	6
Portugal> 40 m	0	0
Scotland< 40 m	17	7
Scotland> 40 m	285	132
Sweden< 40 m	112	11
Sweden> 40 m	71	71
Total	9913	685

Table 6. Strata, numbers of PSU and sample events per strata for scenario 1F. Each country has 2 strata for vessels above and below 40 meters.

Strata: Vessel flag country	Total number of trips	Numbers of planned sampling Events
Germany < 40 m	2	2
Germany > 40 m	32	30
Denmark < 40 m	1566	58
Denmark > 40 m	321	163
England < 40 m	147	5
England > 40 m	48	48
Spain< 40 m	4784	11
Spain> 40 m	1	0
France< 40 m	1541	9
France> 40 m	27	27
Ireland< 40 m	216	15
Ireland> 40 m	173	63
Netherlands < 40 m	0	0
Netherlands > 40 m	39	39
Portugal< 40 m	531	5
Portugal> 40 m	0	0
Scotland< 40 m	17	7
Scotland> 40 m	285	176
Sweden< 40 m	112	5
Sweden> 40 m	71	19
Total	9913	682

Table 7. No strata and a random selection of present numbers of sampling events across countries for scenario 2A.

Sampling unit	Strata	Total number of trips	Numbers of planned sampling events
PSU	Not available	4433	675
SSU	Not available	9913	675

Table 8. Sampling country is strata and a present effort within a country is maintained for scenario 2B.

Sampling unit	Strata: Vessel flag country	Total number of trips	Numbers of planned sampling events
PSU	Germany	58	8
	Denmark	699	240
	Great Britain	289	178
	Spain	1750	11
	France	942	65
	Ireland	148	61
	Netherlands	98	51
	Portugal	245	6
	Sweden	52	52
	Faroe Island	2	0
	Norway	150	0

	Total	4433	672

Sampling unit	Strata	Total number of trips	Numbers of planned sampling events
SSU	Not available	9913	675

Table 9. Sampling country is strata and effort is re-distributed according to landings in scenario 2C.

Sampling unit	Strata: Vessel flag country	Total number of trips	Numbers of achieved sampling events
PSU	Germany	58	37
	Denmark	699	187
	Great Britain	289	115
	Spain	1750	14
	France	942	15
	Ireland	148	65
	Netherlands	98	98
	Portugal	245	6
	Sweden	52	9
	Faroe Island	2	2
	Norway	150	119
	Total	4433	672

Sampling unit	Strata	Total number of trips	Numbers of planned sampling events
SSU	Not available	9913	667

Table 10 Port size is strata and effort is re-distributed according to landings in big/ small ports in scenario 2D.

Sampling unit	Strata	Total number of trips	Numbers of planned sampling events
PSU	Large ports	1108	631
PSU	Small ports	3325	44

Sampling unit	Strata	Total number of trips	Numbers of planned sampling events
SSU	Not available	9913	675

Table 11. Port size and landing country is strata and effort is re-distributed according to landings in big/ small ports for scenario 2E.

Strata: landing country + harbor size	Total number of marked days	Numbers of planned sampling events
Germany large port	53	7
Germany small port	5	4
Denmark large port	564	237
Denmark small port	135	6
Great Brittan large port	175	169
Great Britain small port	114	12
Spain large port	0	0
Spain small port	1750	10
France large port	16	16
France small port	926	41
Ireland large port	80	56
Ireland small port	68	8
Netherlands large port	98	50
Netherlands small port	39	39
Portugal large port	0	0
Portugal small port	245	5
Sweden large port	0	0
Sweden small port	52	52
Norway large port	122	0
Norway small port	28	0
Faroe Island small port	2	0
Faroe Island large port	0	0

Total	4433	673

Table 12. Port size and landing country are strata and effort is re-distributed according to landings by country and in big/ small ports for scenario 2F.

Strata: landing country + harbor size	Total number of marked days	Numbers of planned sampling events
Germany large port	53	33
Germany small port	5	5
Denmark large port	564	183
Denmark small port	135	6
Great Britain large port	175	108
Great Britain small port	114	9
Spain large port	0	0
Spain small port	1750	12
France large port	16	7
France small port	926	9
Ireland large port	80	59
Ireland small port	68	8
Netherlands large port	98	98
Netherlands small port	0	0
Portugal large port	0	0
Portugal small port	245	4
Sweden large port	0	0
Sweden small port	52	7
Norway large port	122	103
Norway small port	28	18
Faroe Island small port	2	2
Faroe Island large port	0	0
Total	4433	671

Table 13. Port size and landing country are strata. Large ports are always sampled, small ports 4 times a year in scenario 2G.

Strata: landing country + harbor size	Total number of marked days	Numbers of planned sampling events
Germany large port	53	53
Germany small port	5	5
Denmark large port	564	564
Denmark small port	135	20
Great Britain large port	175	175
Great Britain small port	114	36
Spain large port	0	0
Spain small port	1750	192
France large port	16	16
France small port	926	120
Ireland large port	80	80
Ireland small port	68	60
Netherlands large port	98	98
Netherlands small port	0	0
Portugal large port	0	0
Portugal small port	245	52
Sweden large port	0	0
Sweden small port	52	28
Norway large port	122	122
Norway small port	28	28
Faroe Island small port	2	2
Faroe Island large port	0	0
Total	4433	1651

Sampling unit	Strata	Total number of trips	Numbers of planned sampling events

SSU	Vessel<40 m	8916	965
SSU	Vessel>40 m	997	982

Table 14. Summary of the 20 different scenarios in the self-sampling design and port sampling design by stock. 1a1-1a4 are different efforts (n) but with the same settings as 1a the same is the case with 1f1-1f3 – different effort (n) but same settings as in 1f.

Scenario	stock	true ton	estimated ton	SE	RSE	MRSE country	MRSE area	MRSE quarter
1a	her-nsea	306419	309224	4003	0.013	0.034	0.04	0.028
1a1	her-nsea	306419	303343	2006	0.007	0.02	0.022	0.015
1a2	her-nsea	306419	306285	5235	0.017	0.046	0.06	0.038
1a3	her-nsea	306419	308005	2457	0.008	0.024	0.028	0.018
1a4	her-nsea	306419	308800	2774	0.009	0.027	0.031	0.02
1b	her-nsea	306419	307340	1842	0.006	0.02	0.034	0.021
1c	her-nsea	306419	308481	2273	0.007	0.018	0.039	0.025
1d	her-nsea	306419	306466	815	0.003	0.01	0.013	0.006
1e	her-nsea	306419	307183	575	0.002	0.009	0.01	0.005
1f	her-nsea	306419	306338	438	0.001	0.006	0.011	0.006
1f1	her-nsea	306419	304577	841	0.003	0.012	0.019	0.01
1f2	her-nsea	306419	306305	98	0	0.002	0.005	0.001
1f3	her-nsea	306419	306487	187	0.001	0.002	0.01	0.002
2a	her-nsea	306419	305621	2532	0.008	0.027	0.028	0.018
2b	her-nsea	306419	286410	1062	0.004	0.017	0.014	0.01
2c	her-nsea	306419	305499	847	0.003	0.012	0.011	0.008
2d	her-nsea	306419	308146	1204	0.004	0.031	0.023	0.008
2e	her-nsea	306419	290244	1062	0.004	0.016	0.015	0.01
2f	her-nsea	306419	306580	800	0.003	0.011	0.017	0.007
2g	her-nsea	306419	306616	115	0	0.001	0.004	0
1a	mac-nea	484969	482718	4086	0.008	0.029	0.059	0.026
1a1	mac-nea	484969	485402	2261	0.005	0.017	0.033	0.014
1a2	mac-nea	484969	483357	5855	0.012	0.04	0.082	0.037
1a3	mac-nea	484969	484365	2755	0.006	0.021	0.041	0.018
1a4	mac-nea	484969	487113	3120	0.006	0.023	0.045	0.02
1b	mac-nea	484969	484238	1883	0.004	0.029	0.047	0.018
1c	mac-nea	484969	487411	1759	0.004	0.024	0.049	0.021
1d	mac-nea	484969	483899	1202	0.002	0.014	0.073	0.008
1e	mac-nea	484969	484163	916	0.002	0.012	0.05	0.009
1f	mac-nea	484969	485294	589	0.001	0.012	0.047	0.009
1f1	mac-nea	484969	486356	1135	0.002	0.023	0.077	0.016
1f2	mac-nea	484969	485299	199	0	0.005	0.025	0.004
1f3	mac-nea	484969	485227	355	0.001	0.01	0.045	0.008
2a	mac-nea	484969	485587	3350	0.007	0.026	0.051	0.02
2b	mac-nea	484969	343048	1364	0.004	0.025	0.035	0.016
2c	mac-nea	484969	483575	1481	0.003	0.022	0.046	0.015
2d	mac-nea	484969	481486	2246	0.005	0.036	0.085	0.02
2e	mac-nea	484969	347932	1116	0.003	0.023	0.038	0.015
2f	mac-nea	484969	480783	1119	0.002	0.028	0.056	0.016
2g	mac-nea	484969	484843	223	0	0.006	0.019	0.003
1a	spr-all	167597	167530	1451	0.009	0.033	0.028	0.017
1a1	spr-all	167597	167723	799	0.005	0.02	0.015	0.01
1a2	spr-all	167597	166309	2069	0.012	0.047	0.038	0.024

1a3	spr-all	167597	167755	983	0.006	0.024	0.018	0.013
1a4	spr-all	167597	167952	1153	0.007	0.027	0.021	0.014
1b	spr-all	167597	167663	911	0.005	0.008	0.017	0.012
1c	spr-all	167597	167582	990	0.006	0.014	0.02	0.013
1d	spr-all	167597	168275	1421	0.008	0.032	0.025	0.035
1e	spr-all	167597	167198	609	0.004	0.012	0.012	0.016
1f	spr-all	167597	166915	647	0.004	0.017	0.013	0.018
1f1	spr-all	167597	166709	1013	0.006	0.026	0.022	0.024
1f2	spr-all	167597	167578	395	0.002	0.006	0.005	0.011
1f3	spr-all	167597	168795	595	0.004	0.013	0.009	0.018
2a	spr-all	167597	170200	1519	0.009	0.027	0.022	0.019
2b	spr-all	167597	166725	822	0.005	0.016	0.012	0.011
2c	spr-all	167597	167768	975	0.006	0.014	0.014	0.013
2d	spr-all	167597	167380	593	0.004	0.029	0.021	0.024
2e	spr-all	167597	168955	692	0.004	0.019	0.012	0.016
2f	spr-all	167597	166666	820	0.005	0.016	0.014	0.018
2g	spr-all	167597	167782	158	0.001	0.004	0.003	0.007

ANNEX 12 WP2 CS2 North Sea demersal fisheries sampling designs

The fishPi project
Regional co-operation in fisheries data collection

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2.3 – Case Study fisheries

Deliverable CS2 - North Sea demersal fisheries sampling designs

Lead: MSS, Scotland

Core Team: Alastair Pout, Liz Clarke (MSS, Scotland)
Ana Ribeiro Santos, Jon Elston (CEFAS, UK England)
Patrik Börjesson (SLU Sweden)
Kirsten Birch Harkannson, Marie Storr-Paulsen (DTU Aqua, Denmark)

Data providers: ILVO (Belgium), DTU-Aqua (Denmark), TI-SF (Germany), IFREMER (France), CEFAS (England, UK), IMARES (Netherlands), MSS (Scotland UK), SLU (Sweden).

Table of Contents

1 North Sea demersal fisheries description

The North Sea demersal fisheries description presented in this report is based on the official landings for 2014 provided by each country participating in the fishPi project.

The North Sea is one of the world's most important fishing grounds, international fishing fleets operate in the southern, central and northern North Sea and target both pelagic and demersal fish stocks. One of the most important fisheries in the North Sea is the mixed demersal fishery, using otter trawls or seines. There are different fisheries operating with these gears: 1) A mixed demersal fishery targeting cod and associated species (mainly haddock and whiting in the Western and Northern North Sea, mainly plaice in the South-eastern North Sea) with trawls and seine nets operates over much of the area indicated above. Of particular importance are the areas off Denmark, around Shetland and adjacent to the Norwegian Deeps. The main countries involved are Scotland, Denmark and Germany. 2) A mixed fishery that is characterised by a greater preponderance of 'groundfish' species targeting in particular anglerfish and megrim. The main area of operation for this fishery is along the shelf edge at depths around 200 m and this fishery is particularly important in Scotland. 3) A fishery for Norway lobster. This fishery is distributed to specific grounds throughout the North Sea and Skagerrak. 4) A mixed fishery taking place in the more southerly parts of the North Sea and centred on the eastern Channel in which whiting and non-quota species are important constituents. This is predominantly a French fishery. 5) A 90-99 mm mesh mixed demersal fishery centred on the Skagerrak and prosecuted by Denmark and Sweden. 6) A directed *Nephrops* fishery in the Skagerrak with sorting grid (70-89 mm mesh size) is prosecuted by Swedish vessels (Pastoors and Quirijns, 2014; Rogers and Stocks (2001)).

Other important fisheries in the North Sea include fisheries using beam trawls. The most important beam trawl fishery is the mixed flatfish fishery for sole, plaice and other flatfish species, located in the Southern North Sea and into the Eastern Channel and is operated mainly by the Netherlands, Belgium and Germany (Pastoors and Quirijns, 2014; Rogers and Stocks (2001)).

A number of fixed gears are employed in the North Sea, the most important being gill nets and trammel nets. The most gillnet activity is within a Danish fishery targeting mainly cod and plaice. The importance of anglerfish in this fishery has risen in recent years and activity directed at this species has increased by Scottish flagged vessels. Trammel net fisheries are operated by a number of countries and are particularly important in more coastal waters, for example off the English and French Southern North Sea and Channel coasts for sole. Catches of plaice and cod in these fisheries are also important. There are also fairly small scale fisheries using longlines catch cod, hake and ling.

Most countries also have inshore fisheries prosecuted by under 10m vessels using a variety of passive and towed gears catching a variety of fish, and pots and dredges catching shellfish species.

According to the official landings data provided by each country for 2014, the main demersal fleets operating in the North Sea were DTS (Demersal trawlers and/or demersal seiners), DFN (Drift and/or fixed netters and to a lesser extent the HOK (vessels using hooks), with 48,298, 46,847 and 5,859 registered fishing trips, respectively (Table 1 and Fig. 1). The countries with the highest number of fishing trips were France, England and Denmark (Table 1 and Fig. 1).

Table 7. Total number of trips fishing in the North Sea in 2014, by country and fleet segment.

VESSEL TYPE	BEL	DEU	DNK	FRA	GBE	GBN	GBW	NLD	SCT	SWE	TOTAL NO. TRIPS
DFN	73	170	5433	19364	19113	0	97	2351	51	195	46847
DRB	6	0	0	0	0	0	0	0	0	0	6
DTS	537	568	11219	13869	9218	233	7	1936	7666	3045	48298
FPO	0	0	26	0	0	0	0	94	13	0	133
HOK	0	0	128	3148	668	0	0	1487	358	70	5859
MGO	0	0	0	0	0	0	0	1	0	0	1
MGP	0	0	2	0	0	0	0	0	0	0	2
PG	0	0	0	0	0	0	0	55	0	0	55
PGO	0	0	0	0	1386	0	0	21	0	0	1407
PGP	0	0	0	680	12	0	0	0	0	11	703
PMP	0	0	2	1474	345	0	0	0	0	8	1829
PS	0	0	0	0	0	0	0	0	0	4	4
TBB	2673	258	83	0	940	0	0	4666	45	0	8665
TM	0	0	4	0	0	0	0	37	0	0	41

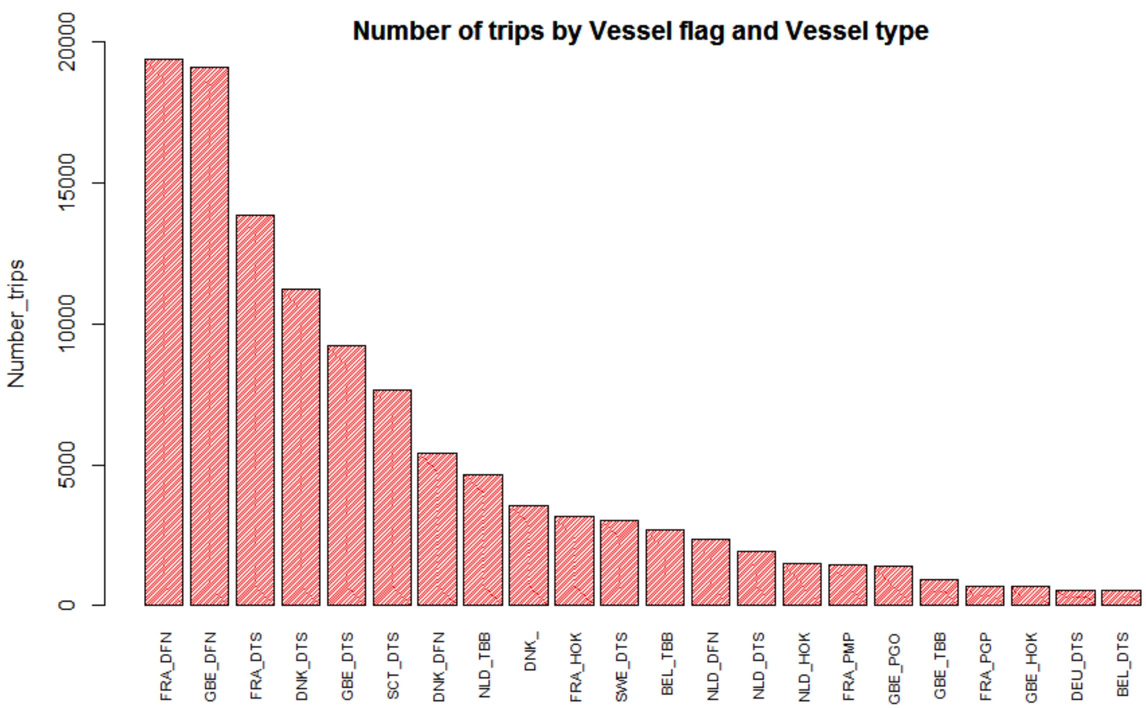


Figure 2 Total number of trips by vessel flag and vessel type combination in 2014, in the North Sea.

The countries with the highest number of demersal vessels fishing in the North Sea are England (GBE), France and Denmark (Table 2 and Fig. 2). Most of these vessels are under 10m length netters and demersal trawlers. These vessels are an important component of the fleet, mainly operating in the Southern North Sea. Despite the high number of small vessels (<12m), their landings account for a small percentage of the landings from the North Sea.

Table 8: Number of vessels by vessel flag and vessel type.

VESSEL TYPE	BEL	DEU	DNK	FRA	GBE	GBN	GBW	NLD	SCT	SWE
DFN	2	9	117	150	442	0	2	83	6	7
DRB	1	0	0	0	0	0	0	0	0	0
DTS	16	20	217	217	220	29	3	52	256	90
FPO	0	0	1	0	0	0	0	10	4	0
HOK	0	0	6	48	29	0	0	73	35	7
MGO	0	0	0	0	0	0	0	1	0	0
MGP	0	0	1	0	0	0	0	0	0	0
PG	0	0	0	0	0	0	0	3	0	0
PGO	0	0	0	0	44	0	0	5	0	0
PGP	0	0	0	8	3	0	0	0	0	1
PMP	0	0	1	46	13	0	0	0	0	1
PS	0	0	0	0	0	0	0	0	0	1
TBB	56	17	2	0	30	0	0	129	1	0
TM	0	0	1	0	0	0	0	8	0	0

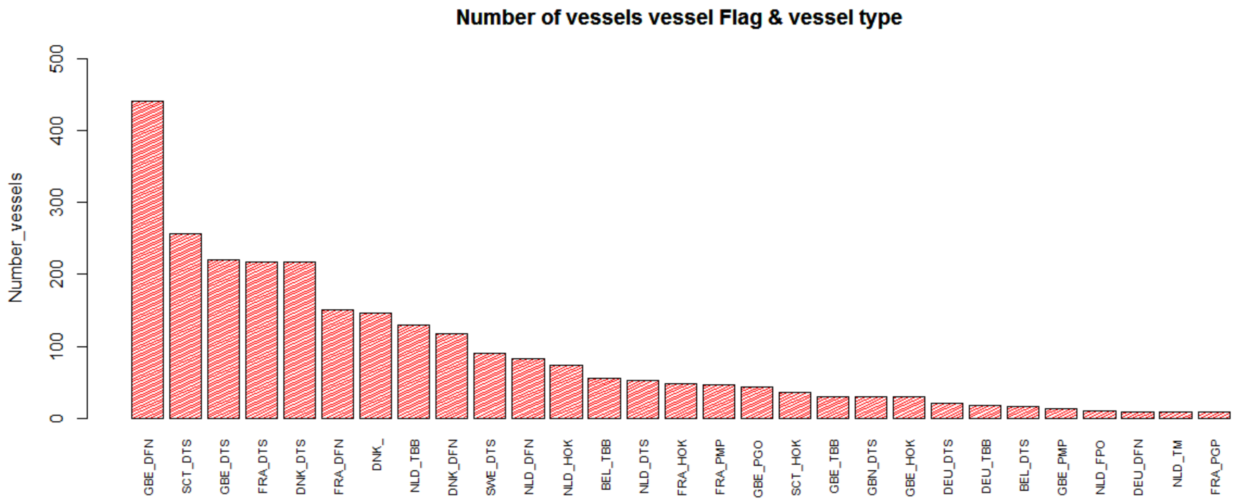


Figure 3: Number of vessels by vessel flag, type and overall length fishing in the North Sea, in 2014.

There were 185 species classified by the International Standard Statistical Classification of Aquatic Animals and Plants (ISSCAAP) as demersal species landed in the North Sea area, in 2014. The total landings of these species were 359,684 tonnes. The species with the most substantial landings were European plaice (*Pleuronectes platessa*), haddock (*Melanogrammus aeglefinus*), cod (*Gadus morhua*), Pollack (*Pollachius virens*), hake (*Merluccius merllucius*), sole (*Solea solea*) and others (Fig. 3). These species account for 65% of the total landed weight.

The countries and fleets that contributed most for the total landings were the Scottish, Danish and French demersal otters and the Dutch beam trawlers, with most of the vessels being bigger than 24m (Fig. 4).

The landings into 38 harbours collectively account for tonnes, equivalent to 95% of the total landed weight for demersal taxa. The main landing ports are Peterhead (Scotland), Hanstholm (Denmark), Harlingen (the Netherlands), Thyborøn (Denmark) and Boulogne-sur-Mer (France) (Table 3).

The predominant recorded métiers in the North Sea demersal fisheries are the otter trawlers, using codend mesh >=120mm (OTB_DEF_>=120) and the beam trawlers (TBB_DEF), accounting for 50k and 49k tonnes of the landings in the North Sea, respectively (Table 4). While the bulk of the landings from the beam trawlers are originated from the Southern North Sea, the otter trawlers landings are more widely spread, but most of the effort is from the central and northern North Sea. (Fig. 5).

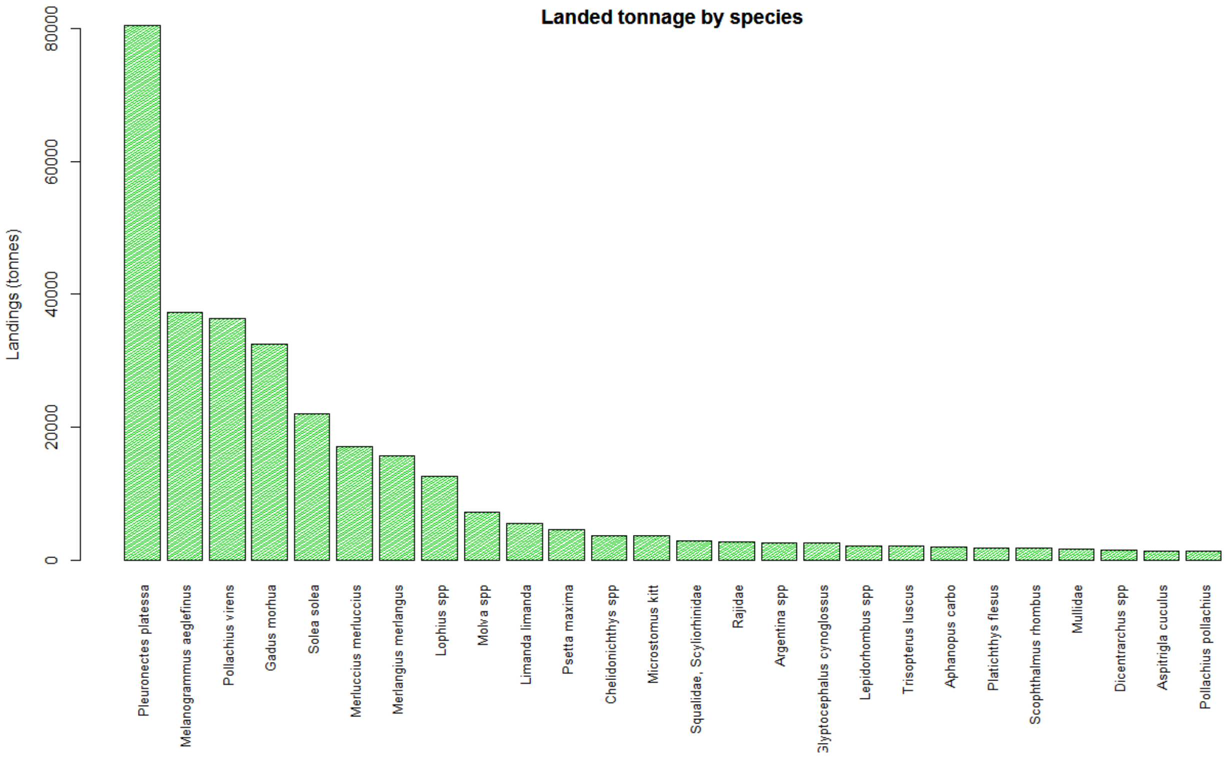


Figure 4: Total landings (tonnes) by species.

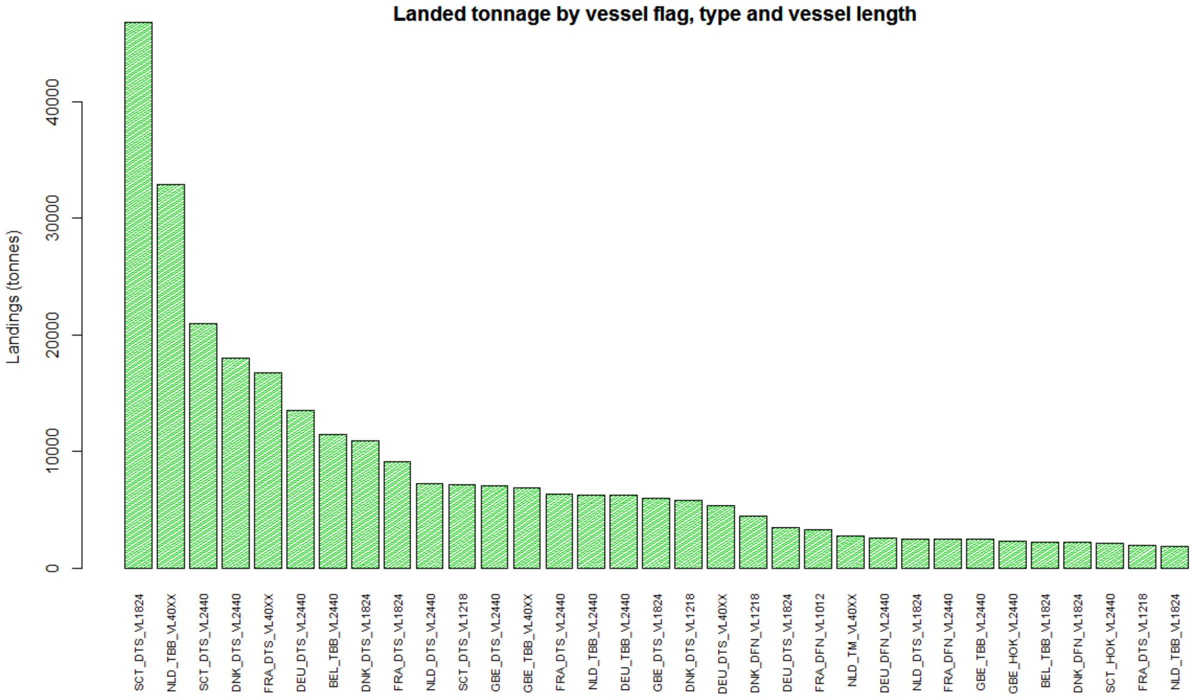


Figure 5: Landings by country, fleet and vessel length

Table 9: Main landing ports in North Sea. The port codes are the United Nations Code for Trade and Transport Locations (UN/LOCODE) five digit port codes.

Country	Port code	Port name	Landings (tonnes)
SCT	GBPHD	Peterhead	43003
DNK	DKHAN	Hanstholm	39179
NLD	NLHAR	Harlingen	26709
DNK	DKTHN	Thyborøn	19730
FRA	FRBOL	Boulogne-sur-Mer	16315
NLD	NLDHR	Den Helder	14214
SCT	GBSCR	Scrabster	13006
SCT	GBLOV	Lochinver	12067
NLD	NLIJM	IJmuiden/Velsen	10802
SCT	GBLER	Lerwick	9896
NLD	NLVLI	Vlissingen	8391
DNK	DKHIR	Hirtshals	8015
SCT	GBFRB	Fraserburgh	6079
SCT	GBKBE	Kinlochbervie	5656
NLD	NLEEM	Eemshaven	5631
DNK	DKHVS	Hvide Sande	5497
NLD	NLSTD	Stellendam	5137
SCT	GBSWY	Scalloway	4631
SCT	GBULL	Ullapool	4154
BEL	BEOST	Oostende (Ostend)	3738
DNK	DKTMD	Thorsminde	3417
FRA	FRRTB	Port-en-Bessin	3320
NLD	NLSCE	Scheveningen	3030
BEL	BEZEE	Zeebrugge	2853
DNK	DKSKA	Skagen	2435
FRA	FRLEH	Le Havre	1786
SCT	GBYFR	Yell and Fetlar	1740
NLD	NLLAN	Lauwersoog	1524
FRA	FRLRT	Lorient	1479
ESP	ESLCG	La Coruña	1477
DNK	DKTHP	Thorup Strand	1473
FRA	FRDPE	Dieppe	1355
FRA	FRFEC	Fécamp	1295
FRA	FRDKK	Dunkerque	1178
FRA	FRCER	Cherbourg	1025
FRA	FRLTR	Le Tréport	993
FRA	FRCOC	Concarneau	916
SCT	GBMLG	Mallaig	878
DEU	DECUX	CUXHAVEN	875
NLD	NLBRS	Breskens	840
NLD	NLDZL	Delfzijl	809

Table 10: Total landings in the North Sea by métier, in 2014.

Métier (Level 6)	Landings (Tonnes)
OTB_DEF_>=120_0_0	50,068
TBB_DEF_70-99_0_0	48,603
PTB_DEF_>=120_0_0	20,934
SSC_DEF_>=120_0_0	19,680
OTB_MCD_>=120_0_0	19,086
OTB_DEF_100_119_0	12,102
TBB_DEF_>=120_0_0	10,771
OTB_DEF_70_99_0	10,092
OTT_DEF_>=120_0_0	9,744
GNS_DEF_120-219_0_0	9,424
SDN_DEF_>=120_0_0	8,250
LLS_DEF_0_0_0	6,809
OTB_DEF_100-119_0_0	6,758
TBB_DEF_100-119	4,989
OTB_DEF_>=120	4,873
OTB_MCD_90-119_0_0	4,785
SSC_DEF_70-99_0_0	4,567
GNS_DEF_>=220_0_0	3,714
OTB_DEF_>=120_0	3,672
TBB_DEF_70-99	3,515
OTB_DWS_>=120_0	3,214
OTM_SPF_32-69_0_0	3,062
GTR_DEF_90_99_0	3,060
OTB_CEP_70_99_0	2,877
TBB_DEF_100-119_0_0	2,650
OTB_CRU_70-99	2,284
OTB_DEF_70-99_0_0	1,934

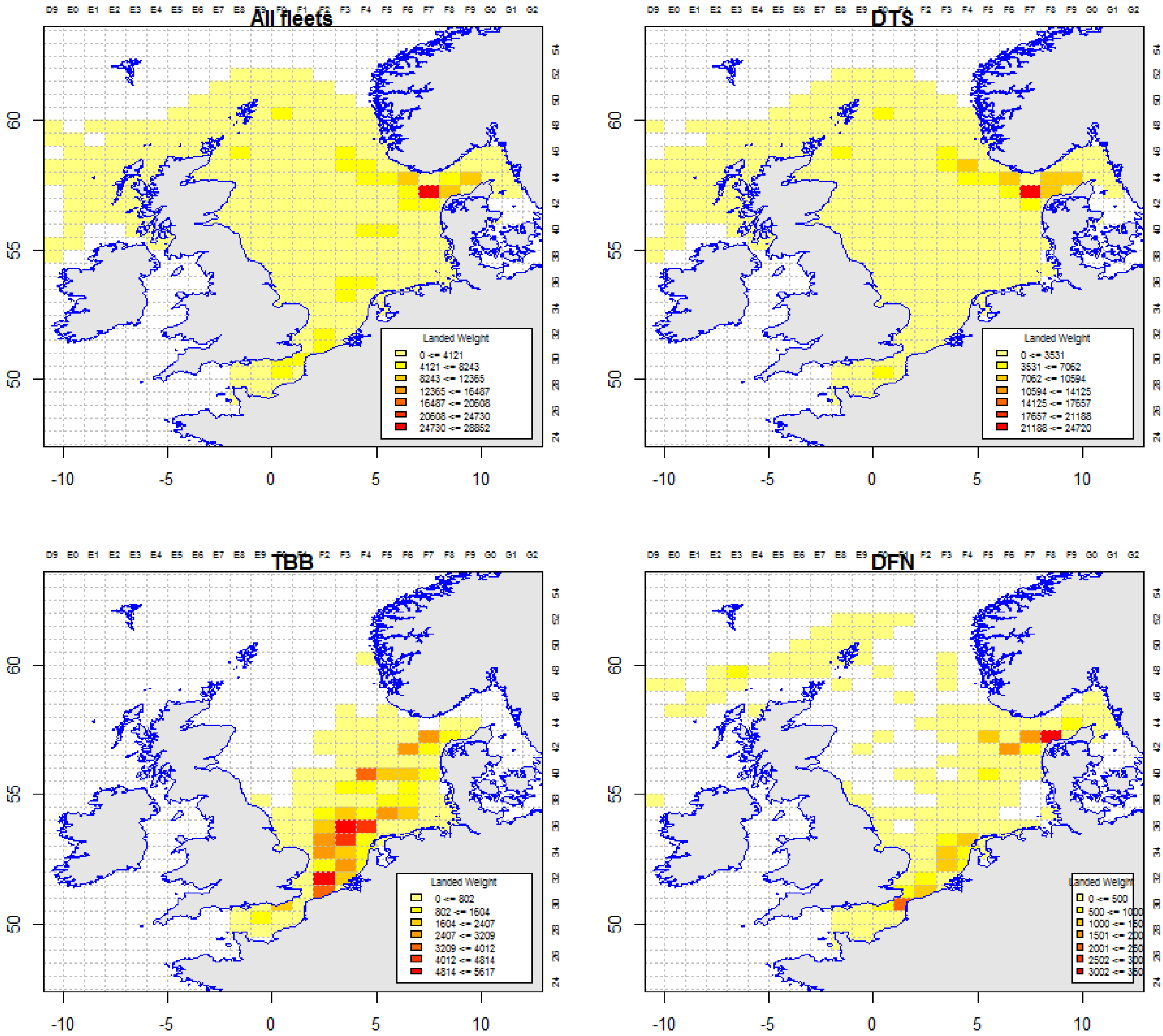


Figure 6: Spatial distribution of landings, by ICES rectangle in the overall fleets and by fleet: DTS (Demersal trawlers and seiners), TBB (beam trawlers) and DFN (Netters).

2 The North Sea demersal data set

Data call

The data call was addressed to each institute involved in the fishPi project with significant fishing fleet activity in the North Sea; data being provided for Belgium, Germany, Denmark, France, England, Northern Ireland, Wales, Netherlands, Scotland and Sweden. The data call requested trip-level landings and sales information for all fish species and *Nephrops* from all vessels using demersal fishing techniques (gillnets, hooks and lines, demersal trawls and demersal seines, beam trawl), with an exemption for dredgers (DRB), vessels using pots and traps (FPO) and pelagic trawlers (TM). The areas of interest were ICES divisions IIIa, IV, VI and VIId. Although area VI sits outside the North Sea region the stock boundaries for a number of important commercial species (anglerfish, megrim, haddock, saithe and hake) cover both areas.

Data cleaning

After data submission by each institute, data were checked and standardised across institutes and collated in one large dataset to be used in simulations. For case study 2 – North Sea demersal fisheries, the following fields were corrected, filtered or added to provide a final target population for the simulations:

- Vessel type (vslType) was standardised to match with the Data Collection Framework (DCF) fleet definition[1] for all countries;
- FAO areas were limited to 27.3.a, , 27.4.a, 27.4.b, 27.4.c, 27.6.a, 27.6.b and 27.7.d;
- Trips using dredges and pots and traps were excluded. Vessel type was based on predominant gear for the reference period so if that vessel used other gears as well as dredges, pots and traps within the reference period they are included in the data set;
- Additional species fields were added to account for industry 'misnomers' and multiple species names for the same species group;
- Data were filtered to only include demersal fish species, using ISSCAAP definitions[2]:
  31 - Flounders, halibuts, soles
  32 - Cods, hakes, haddocks
  33 - Miscellaneous coastal fish
  34 - Miscellaneous demersal fish
  35 - Herrings, sardines anchovies
  38 - Skates and rays chimaeras

Preliminary exploration of the data showed that the species with the highest landings were sandeels (Ammodytidae) from industrial fisheries and it was agreed to remove this species from the dataset because they would be included in the pelagic fisheries.

In order to better distinguish Scottish and English ports, an extra field was created (ctryCode2) based on the port geographical coordinates. The data cleaning and analyses were performed using the software R 3.2.1.

[1] http://datacollection.jrc.ec.europa.eu/wordef/fleet-segment-dcf
[2] http://www.fao.org/fishery/collection/asfis/en

3 On-shore designs for North Sea demersal fisheries sampling

fishPi Case Study 2 core team (Alastair Pout, Liz Clarke, Ana Ribeiro Santos, Jon Elson, Patrik Börjesson)

3.1 Introduction

Estimates of catch weight and numbers at age or length for fisheries in the North Sea region are at present based on data collected under nationally devised sampling programmes by the different countries with fishing fleets operating in the North Sea region. The design of these national sampling programmes reflects the priorities, commitments, logistics and budgetary considerations that operate within the scientific institutes and national administrations of individual nations. In particular for the demersal species, which are predominantly caught in mixed fisheries operated by a number of different national fleets, such an approach to sampling design is likely to be far from optimal. Despite attempts at harmonisation of regional data collection, through such fora as the Regional Co-ordination Meetings (RCM's), the extent to which national sampling designs can be retrospectively coordinated and optimised is limited.

The rationale of a regional design is, from the outset, to provide the best statistical estimates, in terms of the bias and precision, of the catch of the fish species of the region. Here we explore the potential for a single regional (i.e. international) sampling design, designed from the outset to collect data from all the different nations landing demersal species from the North Sea region. This exercise is based on the simulation of the on-shore sampling of individual fishing trips, and the species landed from those trips. In this way a number of alternative sampling designs, that differ in their stratification and effort allocation, can be compared using objective measures of performance. The feasibility of implementing a regional sampling design, and the implications for different nations are considered.

The data set we use has been compiled from the logbook and sales note data provided by all the nations with significant fisheries in the North Sea region. Such a data set is unprecedented in its resolution and scope, and demonstrates the level of cooperation that has been achieved among member states within the region.

3.2 Methods

The Data

The data used for the simulations was based on collated logbook and sales note data provided by eight countries (Belgium, Germany, Denmark, France, England (inc Wales), Netherlands, Scotland and Sweden) involved in data collection under the DCF data collection regulation. These data collectively account for the majority of the demersal landings from the North Sea and North Atlantic area and include small quantities of landings into Spain, Norway, and Ireland. The working data set used for on-shore sampling designs was refined according to area and vessel type and split by year; 2013 and 2014 data being available. The 2013 data consisted of 27,811 site by day combinations, and 116,888 voyages. Here we define a site by day combination as a unique landing location and the unique day of the year on which at least one voyage offloaded the landed fraction of its catch. We define a voyage as a recorded landing by a fishing vessel into a landing location; obviously more than one voyage can land to a landing location on a particular day, and individual vessels make multiple voyages over the course of the year. For the individual voyages the species landed weight (in kg) was available for all the recorded landed species. Landing location was classified using UN/LOCODE five digit location codes. Species codes, landing location names, dates, vessel identification codes were standardised using the criteria and methods set out in section 2. The data set comprised all trips from ICES divisions IVa, IVb, IVc, VIId and IIIa, VIa and VIb for 2013.

Simulations

All the sampling design simulations were based on two-stage sampling involving, firstly, the selection of the arrival location (site) and date (day), (the primary sampling unit (PSU)), and secondly, the selection of voyages from each of the given locations and dates (the secondary sampling (SSU)). The data recorded from each of the selected voyages were the landed weight of all landed species, along with other pertinent details of the vessel and the voyage such as the métier, fishing location etc. This equates, to a full concurrent sample, in the terminology of the DCF sampling regulations. For the purposes of the simulations, we use the landed weight as a proxy for the age distributions and length distributions that would in reality be collected. This is a first step towards the development of sampling designs and the validity of this assumption is one area where further analysis would be merited.

Alternative designs

In total five designs with differing sampling stratifications were evaluated. The sampling stratifications were based on countries and major and minor ports, as detailed in the design descriptions. The total allowable sampling effort was based on the 2015 sampling effort employed by the sampling institutions in the region, which collectively amounts to 947 on-shore visits. The scenarios compared maintain the existing effort regimes across the whole region, although the distribution of sampling effort between sampling nations and institutions could change. The number of secondary sampling units selected was fixed at a maximum of two voyages per site day visit. This was the best whole number that most closely approximated the achieved numbers of sampled voyages of the various national sampling programmes.

A further five scenarios explored the effect of changes to the overall sampling effort. Here the total sampling effort was allowed to vary while maintaining the proportion of sampling visits by country and port strata according to the effort allocation rules and stratification of the preferred design. These overall effort allocations allowed incremental effort increases from 503 to 1359 sampling visits to illustrate potential changes in precision that could be achieved with more or less sampling.

Finally the preferred design was used on the 2014 data set to generate a single realisation of one year's sampling. This served to demonstrate the robustness of the design to changes between years and illustrate the nature of the estimates obtained.

The preferred sampling design is set out in terms of the sampling frame of landing ports, and the number of on-shore visits, by landing country.

The sampling designs tested were as follows:

1. Simple random sampling

No stratification, so site and day were selected with random sampling without replacement from all possible site and day units available. The secondary sampling was then a random selection of up to two voyages, again selected without replacement, from the total number of potential voyages available for selection on the site and day.

2. Six country sampling

The fishing trips in the data set used in the simulations landed into eleven countries, the number of fishing trips landing into each country varying from 23 in the case of Spain, to 9098 for GBR (England and Wales) (Table 1). However, of the eight countries that contributed to the data set, two, Belgium and Germany, presently don't undertake on-shore sampling programmes of fishing trips operating in the North Sea. In addition there were small amounts of landings into Spain, Ireland and Norway. The six country scenario therefore aimed to mimic as much as possible the situation at present by using an effort regime limited to the six countries where on-shore sampling is presently undertaken. The countries with on-shore sampling in this scenario were Denmark, Scotland, England, Netherlands, Sweden and France. The existing sampling levels were assumed for these countries and zero sampling effort was assumed for all other landing countries (Table 2). Total sampling effort was 939 site day visits. The sites and dates were selected at random within the landing country.

Table 1: The distribution of unique site and day primarily sampling units by landing country, these formed the stratum totals for the landing country stratification scenario.

Present Site day distribution by Country										
BEL	DEU	DNK	ESP	FRA	GBR	IRL	NLD	NOR	SCT	SWE
960	89	2939	23	6627	9098	48	2937	42	2940	1692

Table 2: Present distribution of sampling effort by sampling country. Belgium and Germany are allocated zero sampling effort because they presently do not operate on-shore sampling schemes, Ireland, Norway and Spain have been allocated zero sampling effort because they did not submit data to the data set thus they have minimal landing totals. The total sampling effort for the region amounts to 949 visits. In the 10 country design where a country does not currently have an on-shore sampling programme (or did not contribute to the data set) a nominal figure of 2 PSU samples was allocated.

Present Effort allocation by Country										
BEL	DEU	DNK	ESP	FRA	GBR	IRL	NLD	NOR	SCT	SWE
0	0	84		44	339		101		291	80

3. Ten country sampling

This sampling stratification assumed that all the countries receiving landings from the area would participate in on-shore sampling, and were thus sampling strata. The effort allocation followed that of the present situation (Table 2) except that the countries that presently had no on-shore sampling were allocated the minimal on-shore sampling level of 2 visits. This scenario thus represents a situation with sampling effort similar to that at present but with complete coverage across all landing countries albeit at a minimal level. The total effort allocation in this scenario was 949 visits.

4. Major and minor ports

The major and minor ports design was a stratified design with two regional strata, major ports and minor ports, which ignored the landing country where the port is situated when selecting sampling units. The distinction between major and minor ports was made on the basis of the landed weight. The ports that collectively accounted for 95% of the landed weight of the combined demersal species were grouped into the "major" ports stratum. In total 43 ports in eight countries fell into this criterion. All other landing locations, of which there were 321, were grouped into the "minor" ports stratum. The effort allocated to these strata was set at 760 visits (80%) for the major ports and 189 visits (20%) for the minor ports; this gave the same total visit allocation of 949 visits as used for the 10 country scenario but with an allocation that aimed to concentrate 80% of the sampling effort in the major ports. So, while the sampling stratification was based on landed weight, the minor port stratum received relatively greater sampling effort than would be the case were effort to be allocated in direct proportion to landed weight.

5. Port and country

This design linked the country with the major and minor port stratification, with each landing location was allocated to both a landing country and to one of the regional major or minor port groups. This resulted in a design with 19 strata, with the number of major and minor ports by country shown in Table 3. The effort allocation within this design was distributed by sharing the total regional effort in the major and minor port stratification (760:189 site day visits) between countries in proportion to the number of ports in each country within that stratum (Table 4).

The 43 major ports were distributed mainly in France (12), Scotland (10), the Netherlands (8) and Denmark (7). Norway and Sweden had no ports in the major port stratum. The minor ports accounted for the remaining landing locations. They were distributed between countries as shown in Table 3, though 112 of the 324, over a third, were situated in England.

Table 3: The total numbers of ports by port stratification and country. In total there are 364 ports in the regional data set.

	BEL	DEU	DNK	ESP	FRA	GBR	IRL	NLD	NOR	SCT	SWE
Major	2	1	7	1	12	1	1	8	0	10	0
Minor	2	15	27	3	52	112	2	22	8	40	40

Table 4: The effort allocation by major, minor port stratification and country.

	BEL	DEU	DNK	ESP	FRA	GBR	IRL	NLD	NOR	SCT	SWE
Major	35	18	124	0	212	18	18	142	0	177	0
Minor	2	9	16	2	31	66	2	13	5	24	24

Estimation

For each design scenario, 500 simulations were run, each of those simulations being used to generate estimates of the total landed weights, by species and by country, for the 2013 data set. The estimator used was the Horvitz-Thompson estimator (Horvitz & Thompson, 1952), applied here using the R package “survey” (Lumley 2014) with the svyDesign function being used to specify the design, the svyTotal function being used to provided estimates of the total landed weight, and the svyBy function being used to generate estimates by species and country domains.

Evaluation and Comparison

Thus for each scenario, 500 separate estimates were generated for the 1) total landed weight of all species, 2) the total landed weight of each of the 120 species in the data set and, 3) the total landed weight by landing country. The true values of each of these parameters were of course known from the complete data set, for example the true landed weight of all species was 359,684 tonnes, of cod was 29,976 tonnes, of haddock was 41,501 tonnes, the total landed weight of all species into the Netherlands was 84,604 tonnes, the landed weight of all species into Scotland was 120,004 tonnes, etc. Such a situation is akin to having 500 years of sampling data and their associated estimates, each year differing due to the stochastic nature of the sampling.

The percentage deviation of the total landed weight estimates from the true total was calculated as the mean of the 500 estimates minus the true total, divided by the true total. The relative standard error RSE (the standard deviation of the estimate divided by the estimate) provided a scaled measure of the precision of an estimate. The mean RSE was calculated for the 120 species in the data set to give an indication of the precision possible on average for species, and a mean RSE was calculated for the landed total of all species by country to give an indication of the precision possible on average by landing country. The design effect, D_{eff}, is the variance of the estimator for a design relative to the variance achieved for a random sample (Kish 1965). The design effect was calculated using the mean variances of the estimates of the 120 species in the data. The effective sample size provides an intuitive measure of the merits of a design by scaling the actual sample size of a random sample by the design effect. The effective sample size was estimated assuming a 950 visit design.

$$Design\ effect\ (D_{eff}) = \frac{s^2_{design}}{s^2_{srs}}$$

$$Effective\ sample\ size = \frac{n}{D_{eff}}$$

Designs were therefore assessed on 1) their overall deviation in the estimate of the population total, 2) their RSE over species and 3) their RSE over countries, 4) Design effect, 5) Effective sample size. Collectively these measures were used to ascertain which sampling design proved the most effective. The simple random scenario provided a basis on which to gauge the performance of the other designs.

For species estimates in addition to the standard error, the % deviation and the RSE estimates, a 95% confidence interval was generated based on the 0.25 and 97.25 percentiles of the distribution of the estimates. Fifteen species which are assessed in the ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK) and the ICES Working Group on Celtic Seas Ecoregion (WGCSE), were tabulated to show the % deviation in total estimates, RSE and 95% quantile values, and, for a number of these, the distribution of the sample estimates was plotted as a histogram. For all species, a scatter plot of landed weight estimates against the true values, with 95% confidence intervals was plotted, as was the recorded % deviation in relation to the sample size for each species. The sample size for the species was simply the number of recorded instances of the species in the sample, i.e. how many of the sampled voyages had landed the species in question.

3.3 Results

Design comparisons

The percentage deviation and RSE for the five scenarios are shown in Table 5. In terms of the deviation on the total estimate the major-minor ports design performs best followed by the port and country designs. These are both within +/- 1% of the true value. The random sample has a % deviation of 1.252, however as a random sample will by definition be unbiased this is a reflection of the stochastic nature of the sampling. Of more concern are the % deviations in excess of -3% for both the country designs. Although small, these are likely to reflect actual bias in the estimates because of the lack of sampling effort for some landing countries, rather than random deviation due to stochastic sampling error.

The mean relative standard error over species is best for the major and minor ports design, and port and country design, both at 0.81 and 0.82 respectively. The six country design with existing effort allocations performs particularly badly at 1.15 and shows only a slight improvement to 1.15 with the inclusion of four other sampling countries. The random design is intermediate at 1.04.

The mean relative standard error across countries is best with the port and country design at 0.32, a substantial improvement on any of the other designs. The six country design with the existing effort allocations has a RSE of 0.52, while that of the 1 country stratification is 0.60. Despite not including country in the stratification, the port design provides a more precise estimate by country than the six country design. All the stratified designs provide more precise estimates on average for total landed weight by country than does the random design.

The design effect indicates that for the precision of species estimates, the 6 country design performs less well than a simple random sample, all the other designs perform better than a random sample. The port design provides the best design for precision of species estimates. The effective sample size demonstrates the extent to which a port design would utilise the available sampling effort, providing over a more than fourfold increase in samples compared to the country design.

In conclusion the port and country design offers the best combination of unbiased overall estimates and precision levels for species and country. A regional port design also performs substantially better than a simple random design however the logistics of a regional design that does not involve a national element would prove challenging to implement. The country stratification design with the existing sampling allocations performs particularly badly in that the overall estimate shows the most deviation and the precision across species is poor. Increasing coverage with across the countries has only a minimal effect in lessening the deviation and improving the precision.

Figures 1 to 15, show, for each of the designs, the distribution of simulation estimates for some of the main commercial species, a scatter plot of landed weight by species estimates against the true values, and the percentage deviation of the species estimates in relation to the sample size.

Table 5: Comparison of the performance of the five sampling designs. The criteria used are the % deviation of the total estimate, the mean relative standard error over the 120 species in the 2013 data set and the mean relative standard error of the total landed weight by country. The design effect is the ratio of the variance achieved for a design relative to the simple random design, and the effective sample size by design is based on each design having 950 sampling events.

Design	% Deviation of total estimate	mean RSE species	mean RSE country	Species SE	Deff	n/Deff
Simple random	1.252	1.041	0.6384	1550		
6 country	-3.561	1.226	0.5248	1933	1.55	611
10 country	-3.225	1.156	0.6075	1211	0.61	1556
Major and minor ports	0.038	0.816	0.5706	912	0.35	2745
Ports and country	0.712	0.821	0.3191	1055	0.46	2052

Species Estimates

The predictions for individual species of interest to WGNSSK and WGCSE, are shown in Table 6. Most species provide a low % deviation from the known mean values with RSE much lower than the mean RSE across all 120 species in the full data set. This reflects the frequency of the landing events, and thus the number of samples obtained, for these species of commercial interest. Of interest is the finding that the estimates for saithe are particularly variable, which will be a characteristic of the landing practices for this species. The distribution of estimates for saithe shows a particularly skewed positive distribution indicative of occasional very large landings. This demonstrates that it needs to be born in mind that precision is a function both of sample size but also the inherent variability of the data, hence increased sample size to minimise variance for saithe would be both ineffective and a waste of resources.

Table 6. Species landed weight "true" values from the known data set, and the estimated values, standard error, % deviation, relative standard error and 95% confidence interval from 500 simulations using the port and country sampling design.

		True value	Estimate	SE	% Deviation	RSE	95% CI	
Atlantic cod	*Gadus morhua*	29976	30216	5262	0.80	0.20	21994	45080
Haddock	*Melanogrammus aeglefinus*	41501	41235	5906	-0.64	0.14	30620	52795
Whiting	*Merlangius merlangus*	20267	20692	4510	2.10	0.25	12635	32621
European hake	*Merluccius merluccius*	17410	17726	4133	1.81	0.33	10878	29474
Saithe(=Pollock)	*Pollachius virens*	53363	55110	21911	3.27	1.06	25895	123122
Monkfishes nei	*Lophius spp*	13013	12955	2453	-0.44	0.22	9233	19588
Megrims nei	*Lepidorhombus spp*	2463	2439	424	-0.98	0.18	1733	3396
European plaice	*Pleuronectes platessa*	88777	88608	10935	-0.19	0.13	67445	112428
Witch flounder	*Glyptocephalus cynoglossus*	1921	1872	395	-2.53	0.23	1194	2923
Common sole	*Solea solea*	23631	23730	3122	0.42	0.14	17899	30793
Turbot	*Psetta maxima*	4861	4879	867	0.39	0.18	3403	6894
Pollack	*Pollachius pollachius*	1502	1486	356	-1.07	0.28	869	2413
Brill	*Scophthalmus rhombus*	1913	1914	263	0.07	0.14	1443	2551
Lemon sole	*Microstomus kitt*	3926	3921	631	-0.14	0.18	2831	5657
European flounder	*Platichthys flesus*	1806	1840	441	1.85	0.26	1118	3010

Sampling Effort

Adjustments of effort (Table 7) suggest that, as would be expected, measures of precision improve as the total number of primary sampling units increases. The relative gain in precision for species improves from 0.946 with 502 PSU samples to 0.783 with 1340 PSU samples. The gain is however not a smooth function of PSU number due to Monte Carlo error. Precision for the estimates of landed species into countries improves steadily with increased PSU samples.

It is noticeable that the country and port design, even at 502 PSU, has better measures of % deviation and precision than the simple random design, and both the country designs do with sampling effort at ~ 955 PSU.

Table 7. Effort changes and their effect on RSE with the port and country design.

Effort changes with the port country design	% deviation on total estimate	mean RSE species	mean RSE country
502 PSU visits	0.521	0.946	0.432
653 PSU visits	-0.283	0.976	0.350
804 PSU visits	0.257	0.863	0.338
955 PSU visits	0.712	0.821	0.319
1098 PSU visits	0.409	0.851	0.297
1340 PSU visits	0.296	0.783	0.259

Preferred sampling design

The stratification totals by port and country and country only sampling design is set out in Table 8 with the number of major and minor ports by country, and the present and envisaged effort. Effort is in terms of the number of unique visits to a port day combination for the country, the primary sampling unit. In comparison to the present situation the most notable changes are that England, Scotland and Sweden would reduce their on-shore sampling, while France, the Netherlands, Denmark, Belgium, Germany would increase their on-shore sampling. The situation with Ireland and to a lesser extent Norway and Spain is less clear cut because while the design suggests on-shore sampling should increase, these countries did not provide data and therefore existing sampling is not known. The port and country design sampling frames using the UN/LOCODE five digit port codes are shown in Table 9.

Table 8: Port and country design: Distribution of major and minor ports by country and the existing and envisaged PSU sampling effort.

Effort allocation by Country and Size											
	BEL	**DEU**	**DNK**	**ESP**	**FRA**	**GBR**	**IRL**	**NLD**	**NOR**	**SCT**	**SWE**
Major port allocation	35	18	124	0	212	18	18	159	0	177	0
Minor port allocation	2	9	16	2	31	66	2	13	5	24	24
Present effort	0	0	84		44	339		101		291	80
Envisaged new effort	37	27	140	2	243	84	20	172	5	201	24
Change (number of PSU visits)	37	27	56	2	199	-255	20	71	5	-90	-56

Table 9. Port and country design: Sampling frames of major and minor ports by country.

Sampling Frame Regional on-shore North Sea demersal Design.																				
Stratum total N	**2**	**2**	**1**	**15**	**7**	**27**	**12**	**52**	**1**	**111**		**9**	**21**	**10**	**40**	**40**	**8**	**1**	**2**	**4**
Effort Allocation	**35**	**2**	**18**	**9**	**124**	**16**	**212**	**31**	**18**	**66**		**159**	**13**	**177**	**24**	**24**	**5**	**18**	**2**	**2**
	BEL_major	**BEL_minor**	**DEU_major**	**DEU_minor**	**DNK_major**	**DNK_minor**	**FRA_major**	**FRA_minor**	**GBR_major**	**GBR_minor**		**NLD_major**	**NLD_minor**	**SCT_major**	**SCT_minor**	**SWE_minor**	**NOR_minor**	**IRL_major**	**IRL_minor**	**ESP_minor**
	BEZEE	BENIE	DECUX	DESAS	DKHAN	DKSTY	FRCER	FRTBE	GBNSH	GBBLY	GBDVR	NLIJM	NLURK	GBPHD	GBMCD	SERNG	NOSTV	IEKBS	IEROE	ESVGO
	BEOST	BEBBG		DEHHF	DKHIR	DKSTD	FRBOL	FRLVE		GBSCA	GBCLS	NLSCE	NLWRG	GBLER	GBWRY	SEOCO	NOSKU		IECTB	ESCIO
				DEMNK	DKSKA	DKOTB	FRVHH	FRLRH		GBWTB	GBHTH	NLHAR	NLCOL	GBLOV	GBEYM	SESMD	NOEGE			ESLCG
				DELIS	DKHVS	DKRNN	FRRTB	FRRYN		GBHTP	GBBLS	NLEEM	NLBRS	GBMLG	GBKWL	SEGOT	NOHAU			ESPAS
				DEBKE	DKTHN	DKLKK	FRFEC	FRDSM		GBAMB	GBBRT	NLLAN	NLBSE	GBSCR	GBDMB	SEMSD	NOHVA			
				DEBUM	DKTMD	DKEBJ	FRLRT	FRGCP		GBGSY	GBWNZ	NLSTD	NLTEX	GBFRB	GBWHY	SEROO	NO999			
				DEACC	DKTHP	DKROS	FRCOC	FRCM8		GBSUN	GBHYG	NLDHR	NLOHI	GBSWY	GBBUC	SEKLD	NOAES			
				DEFES		DKNRR	FRLEH	FRGFR		GBRER	GBSZB	NLVLI	NLCPD	GBULL	GBSTO	SEGO2	NOFOS			
				DENRD		DKLIL	FRLTR	FROUI		GBRYE	GBMGT	NLDZL	NLNTJ	GBKBE	GBABD	SEELO				
				DENMK		DKRDS	FRDPE	FRTJE		GBMLF	GBBSI		NLSNN	GBYFR	GBFIE	SELYS				
				DEFR1		DKSLS	FRCQF	FR2GX		GBNHV	GBDWL		NLBZM		GBBRR	SEGRE				
				DEBSK		DKHAS	FRDKK	FRSML		GBSHO	GBSWE		NLEAZ		GBBWK	SESLO				
				DENOR		DKALK		FRELS		GBBRX	GBEMS		NLRSL		GBHOP	SEKUN				
				DEGRE		DKODO		FRHON		GBSWA	GBHRZ		NLEUR		GBSKR	SEFIT				
				DENHS		DKESB		FRROS		GBNYL	GBPTL		NLPTT		GBWMD	SEHNO				
						DKTYB		FRPVU		GBFAL	GBSND		NLTHD		GBTHT	SEBOB				
						DKGLE		FRSV3		GBPLY	GBBOC		NLKAT		GBOBA	SEOC2				
						DKSRG		FRBES		GBGTY	GBGWG		NLYSK		GBWIC	SESMO				
						DKVES		FRLQ4		GBCBT	GBCRR		NLHLW		GBRCD	SESKM				
						DKLVG		FRYYO		GBPOO	GBWEX		NLKOG		GBNUI	SEHRO				
						DKBKH		FRQUY		GBCCH	GBEXM		NLTNZ		GBSZT	SEBRL				
						DKLNS		FRBAY		GBBSH	GBBLC				GBBMO	SEMLO				
						DKGRE		FRZJZ		GBLIT	GBDEX				GBGAR	SEFOT				
						DKFDH		FRSBI		GBEBO	GBBCS				GBKYL	SEFIS				
						DKHRG		FRF2N		GBCOW	GBHUL				GBPRT	SEGLN				
						DKBOK		FRGPO		GBHRN	GBPMD				GBPWM	SEHKL				
						DKGED		FRQUI		GBPME	GBBDZ				GBBUH	SEHAV				
								FRV35		GBRMG	GBWLS				GBWRB	SEEDK				
								FRVR4		GBSSY	GBITC				GBARB	SEHYP				
								FRLN5		GBQUB	GBWSM				GBCRA	SETRE				
								FRAA2		GBWTS	GBSSH				GBCSA	SEDON				
								FRVRU		GBHTG	GBPRJ				GBTHR	SEKSS				
								FRISM		GBWMS	GBMVG				GBMON	SEKNI				
								FRCAY		GBALB	GBSHU				GBLSS	SEHVS				
								FRLPA		GBWOT	GBLOE				GBATS	SEBUA				
								FRQRU		GBLOW	GBNBC				GBGDN	SEHUN				
								FRHTR		GBLYM	GBTRN				GBBTL	SEKAR				
								FRLM8		GBFWE	GBBOS				GBBDF	SEHBG				
								FRLM2		GBSOU	GBPAG				GBSCC	SEKKN				
								FRUAT		GBDSS	GBIPS				GBSRH	SEVJR				
								FRGVC		GBFOL	GBRVF									
								FRLSO		GBWEY	GBPAD									
								FRLC5		GBLSH	GBCCS									
								FRCJH		GBHRW	GBNCL									
								FRVS9		GBAPP	GBFLE									
								FRBR5		GBTOR	GBDIN									
								FRCFP		GBSWD	GBTAI									
								FRCRT		GBBDT	GBTOL									
								FRAUD		GBSTZ	GBLYR									
								FROTY		GBLOS	GBCWA									
								FRGRV		GBFAV	GBLWC									
								FRCFR		GBCAN	GBTAB									
										GBMAL	GBWIV									
										GBRFD	GBRCS									
										GBBBP	GBGIR									
										GBFXT										

Predictions for the 2014 data set

The preferred design, ports and country, was used to sample the 2014 version of the data set and generate estimates for the main species of interest shown in Table 10 and plotted in Fig. 16. A common estimate of the variability in the estimates is provided by 95% confidence intervals (CIs), which are calculated as the estimate +/- 1.96 times the standard error of estimate. This gives an indication of the likely range of estimates possible if different samples of the same size had been taken, though the assumption of normality of the estimates made here is unlikely to hold for many species and the true CI would probably be different given the non-normal distribution of estimates for most species. Confidence intervals are expected to contain the true population value 95% of the time. These interval estimates covered the true value in 12 of the 15 species considered (Whiting, Pollack and Flounder being the species with the inaccurate confidence intervals). The deviation of estimates ranged from -50% for Pollack, and -47% for Whiting to 3% for Sole and -3% for Hake. Eight of the fifteen species considered had % deviation values below +/- 10%.

Such a single realisation serves to illustrate the point that regardless of the overall unbiased nature of the estimator used, for any single sample the stochastic nature of the sampling process will generate estimates that for some species are poor predictors of the true value in the population. The magnitude of the deviation in these estimates will be determined in part by the inherent variability in the data being collected as measured by the relative standard errors for the species concerned. It should also be noted that the methods used to generate confidence intervals for such estimates should reflect the true distribution of the estimates (for example, should not assume normality of the estimates.)

Table 10: Species landed tonnages from a single realisation of the port and country design used on the 2014 data set. The true value is the known totals from the complete data set, the estimates are based on 943 site day sampling units at 142 different landing locations in 11 countries; 1591 individual fishing trips were sampled. The % deviation is the difference between the realised estimate and the true value, and the upper and lower estimates are the estimate +/- 1.96 times the standard error of the estimate for the species.

		Landed Weight	Estimate	% Deviation	RSE	Estimate – 1.96 SE	Estimate + 1.96 SE
Atlantic cod	*Gadus morhua*	32519	30413	-6.48	4221	22139	38687
Haddock	*Melanogrammus aeglefinus*	37283	35477	-4.84	4871	25929	45025
Whiting	*Merlangius merlangus*	15704	8357	-46.78	1084	6232	10483
European hake	*Merluccius merluccius*	17125	16604	-3.05	2563	11581	21626
Saithe(=Pollock)	*Pollachius virens*	36312	38670	6.50	10558	17977	59364
Monkfishes nei	*Lophius spp*	12689	11416	-10.03	1246	8973	13858
Megrims nei	*Lepidorhombus spp*	2192	2465	12.49	551	1386	3545
European plaice	*Pleuronectes platessa*	80507	76603	-4.85	9525	57934	95271
Witch flounder	*Glyptocephalus cynoglossus*	2547	2378	-6.62	625	1153	3603
Common sole	*Solea solea*	22000	22636	2.89	3760	15266	30006
Turbot	*Psetta maxima*	4573	3853	-15.75	625	2629	5077
Pollack	*Pollachius pollachius*	1378	685	-50.27	118	454	916
Brill	*Scophthalmus rhombus*	1793	1439	-19.74	214	1019	1858
Lemon sole	*Microstomus kitt*	3688	3970	7.63	864	2276	5663
European flounder	*Platichthys flesus*	1828	1136	-37.85	208	728	1545

3.4 Discussion

These simulations offer an opportunity to consider sampling designs, from a regional perspective and in relation to the existing sampling programmes operated by countries within the region. As such they can be used to investigate hypothetical situations and stimulate discussion of the implications of regional sampling. The conclusions of these simulations, some of the assumptions underlying these simulations, and the major implications of regional sampling are considered below:

Criteria for evaluation of designs

At present one of the main requirements of data collection is to provide estimates for the ICES single and mixed fish stock assessments. Hence this study has used criteria to evaluate the designs that reflect the ability of the design to generate regional estimates and precision measures for individual species, particularly those currently assessed by WGNSSK and WGCSE. In addition to the species estimates, many countries also have an obvious interest in their own ability to generate estimates, and in estimating the landings into their own ports, hence the RSE measures by country were felt to be likely to be of interest to member states. However the use of total landed weight is likely to have focussed the evaluation of the design to be that which is more optimal for the main commercial species and less suited to evaluating some of the ecosystem criteria important as marine strategy frame work directive indicators.

Preferred regional design

In as much as this study has been able to approximate the existing sampling designs, the findings would suggest that considerable improvements can be achieved in sampling design. The present situation can be characterised as one where there are countries with no on-shore scheme, alongside other countries with substantial effort devoted to on-shore schemes. For the former estimates are not possible (though they could be obtained from at-sea sampling), for the latter the present large effort allocation serves to provided deceptively precise national estimates. Collectively these national designs are responsible for regional estimates that are potentially biased due to poor coverage and imprecise due to sub-optimal distribution of sampling effort; a simple random sample would achieve far better results.

Of the alternative designs explored in this study, the port and country design performs best in the regional context. It potentially can generate unbiased and more precise estimates for individual fish species, while providing national estimates that, overall, are more precise that those obtained at present. The major implications are the stratification of landing locations into major and minor ports, and given the distribution of these ports, the substantial reallocation of sampling effort that could be needed between different countries. The feasibility of such would need to be considered both nationally and within the regional administration.

Reallocation of sampling effort across countries

If the country and port design was implemented, and sampling effort could be freely redistributed across countries, then France, Netherlands, Denmark, Belgium and Germany would need to increase their on-shore sampling. For England, Scotland and Sweden there would be a decrease in on-shore sampling commitment. For England in particular there would be a substantial decrease in comparison to the present levels of on-shore sampling. Spain, Ireland and Norway are not relevant to consider in this context because the fleets of these countries were not included in the case study and their existing sampling commitments can't be objectively analysed.

Access to arrival locations

These simulations are based on the assumption that arrival locations used to define the site and day PSU is a location where the landings can actually be sampled. This may not be a valid assumption in all cases. For example, where the arrival locations are smaller transit ports from which landings are transferred to fish auctions by road, the opportunities for sampling may be limited. Where arrival location and fish auction coincide the opportunities for sampling are generally much more prevalent. Hence any actual design would be predicated on the ability to access fish at the sampling locations. Identifying sampling locations using sales note data is one approach, though the complexity of some of the fish handling arrangements made such analysis impossible during this project.

Random sampling of voyages

These designs assume that random selection of voyages is made from all the voyages present on a given site and day visit. For a number of countries this would require a change in working practices from the métier based quota sampling promoted under the current DCF regulations. It would also mean that foreign flagged vessels would be sampled in the landing country (and estimates generated for these factions of the landings). Random samples can be achieved in the on-shore situation; all that is required is that an audit of the all voyages present can be made, and that a protocol for making a random selection from that list can be devised. Standard vessel selection forms printed with random number strings adjustable by samples size are in operation in a number of sampling countries. The difficulties arise in situations where landings cannot be unambiguously ascribed to particular vessels, or where there are logistical problems in accessing some part of the landings. Selection protocols may need to be devised on a case by case basis to suit the working characteristics of individual sampling sites.

Scope for refinement

The estimates generated from the simulations are by their nature broad and overarching; the deviation in estimates of total landed weight of all species, total landed weight by species and total landed weight by country. This was due to the regional perspective of the design and the time considerations of the analysis. The methods and general approach used in these simulation studies do however offer considerable scope for refining particular criteria of importance for a design, and in generating estimates for more precisely defined domains that are deemed important (such as stocks, countries, métier groups, temporal periods etc). The constraints lie only in the ability to define these variables from with the available data set.

Validity of the estimation metric

These simulations used the landed weight as the metric for the evaluation of alternative designs. They considered the sampling hierarchy only as far as two-stage sampling of vessel within site and day. Further it was assumed that once a vessel was selected a complete audit of the landed weight of all species was possible. In reality on-shore sampling would involve the collection of age structures and length measurements from species and size sorted categories of fish. Hence a further consideration with these simulations is the extent to which landed weight serves as a good proxy for the age and length related variables that are actually to objective of the sampling design. Certainly landed weight is considered highly correlated with total numbers of fish and therefore could indeed constitute a key component of the variability in estimates. In addition, time and accessibility constraints in the real sampling situation may limit the extent to which the sampling of all species and the collection of age structures is possible.

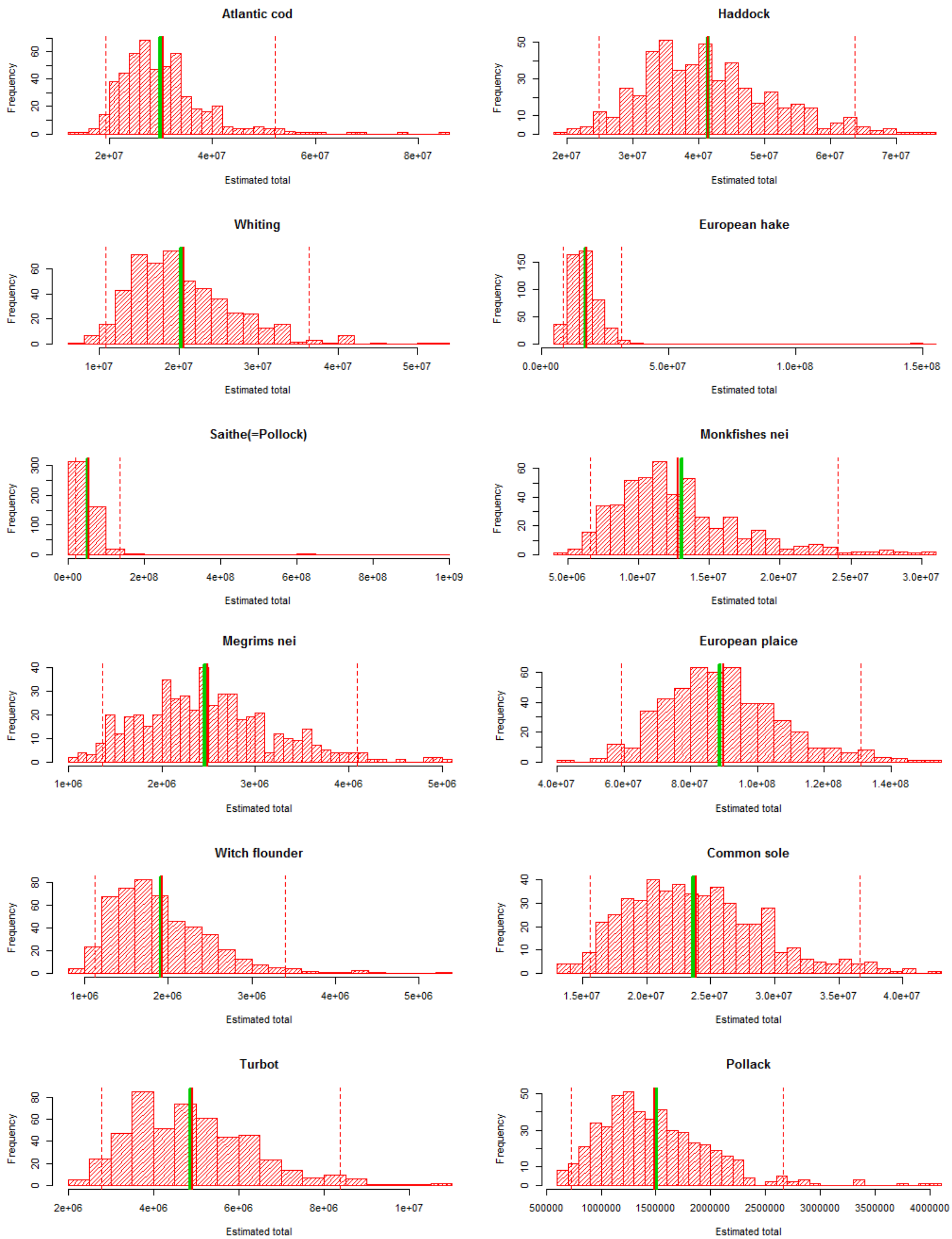


Figure 2. Simple Random Design: distribution of 500 estimates of total landed weight for selected species, the mean and 95% confidence intervals of the estimates are marked as red vertical lines, the true value of the species landed weight in green. It is noticeable that the distribution of landing total estimates has a skewed positive distribution, e.g. for cod, whiting, and particularly hake and saithe.

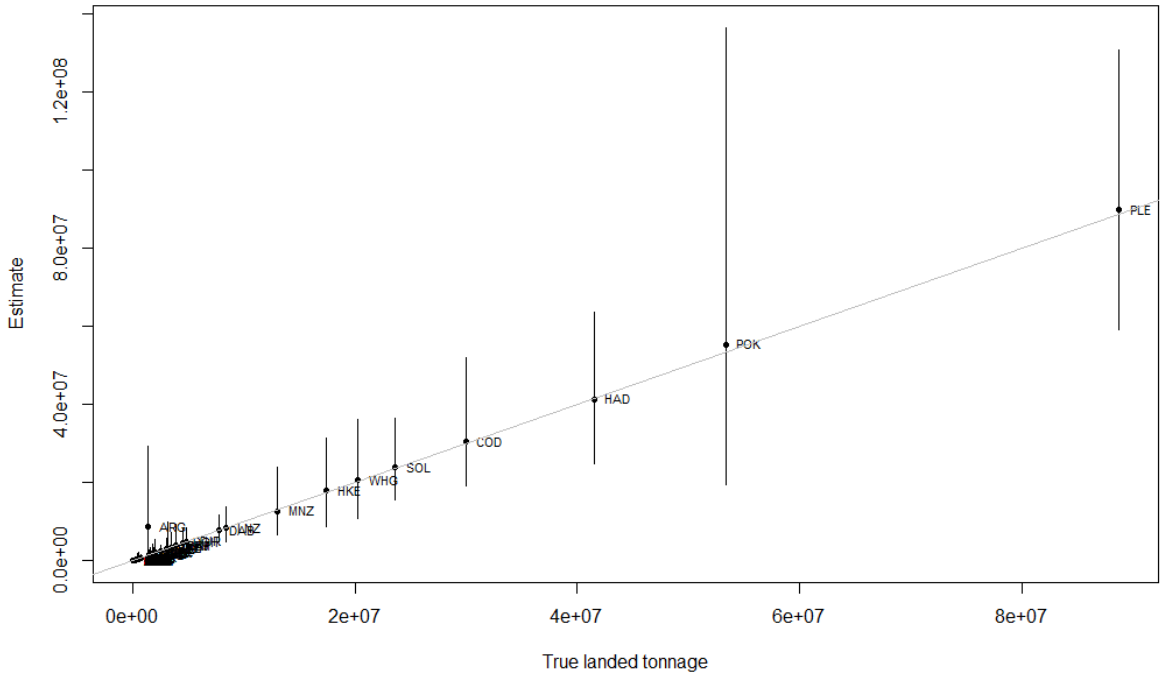


Figure 2: Simple random design: Estimates and 95% confidence intervals of landed tonnages by species. The variability of different species estimates is apparent, as shown by the 95% confidence intervals; saithe (POK) RSE value of 1.81, and argentines (ARG) being two species with particularly high variability. In contrast, haddock (HAD), sole (SOL), and plaice (PLE) have low variability, with RSE values of 0.14 and 0.13.

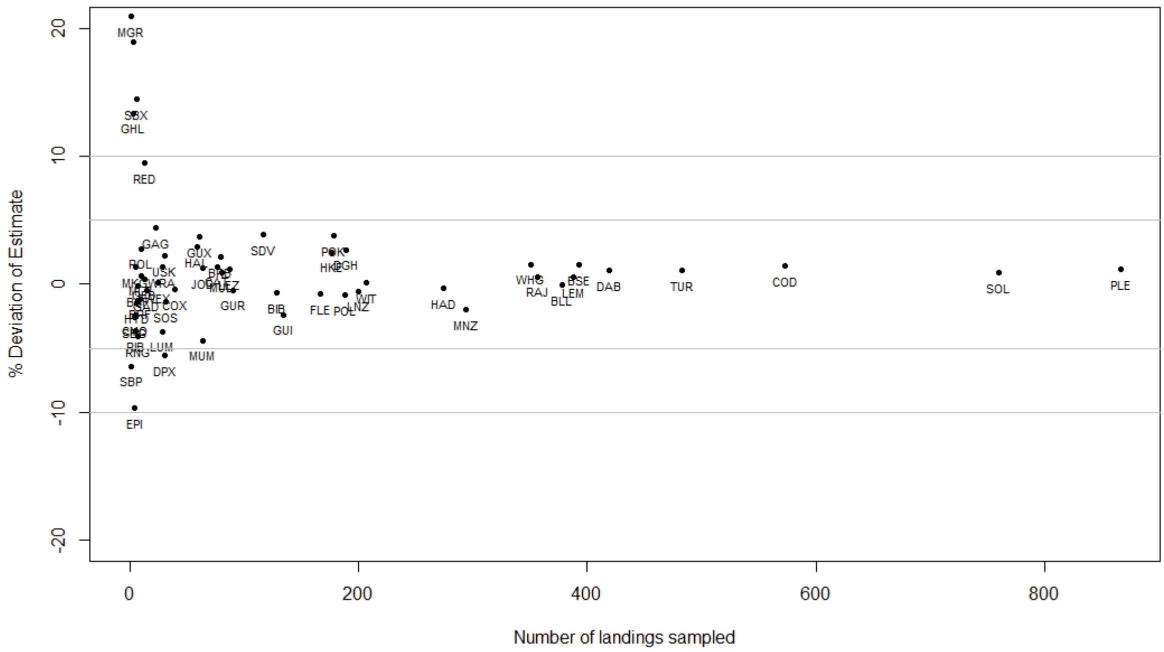


Figure 3: Simple random design: Percentage deviation of the estimate in relation to the number of samples. A simple random sample will generate unbiased estimates for all species, however with small samples sizes i.e. where there are few landings of a particular species, the estimates are more varied and take more simulations to approach the true value.

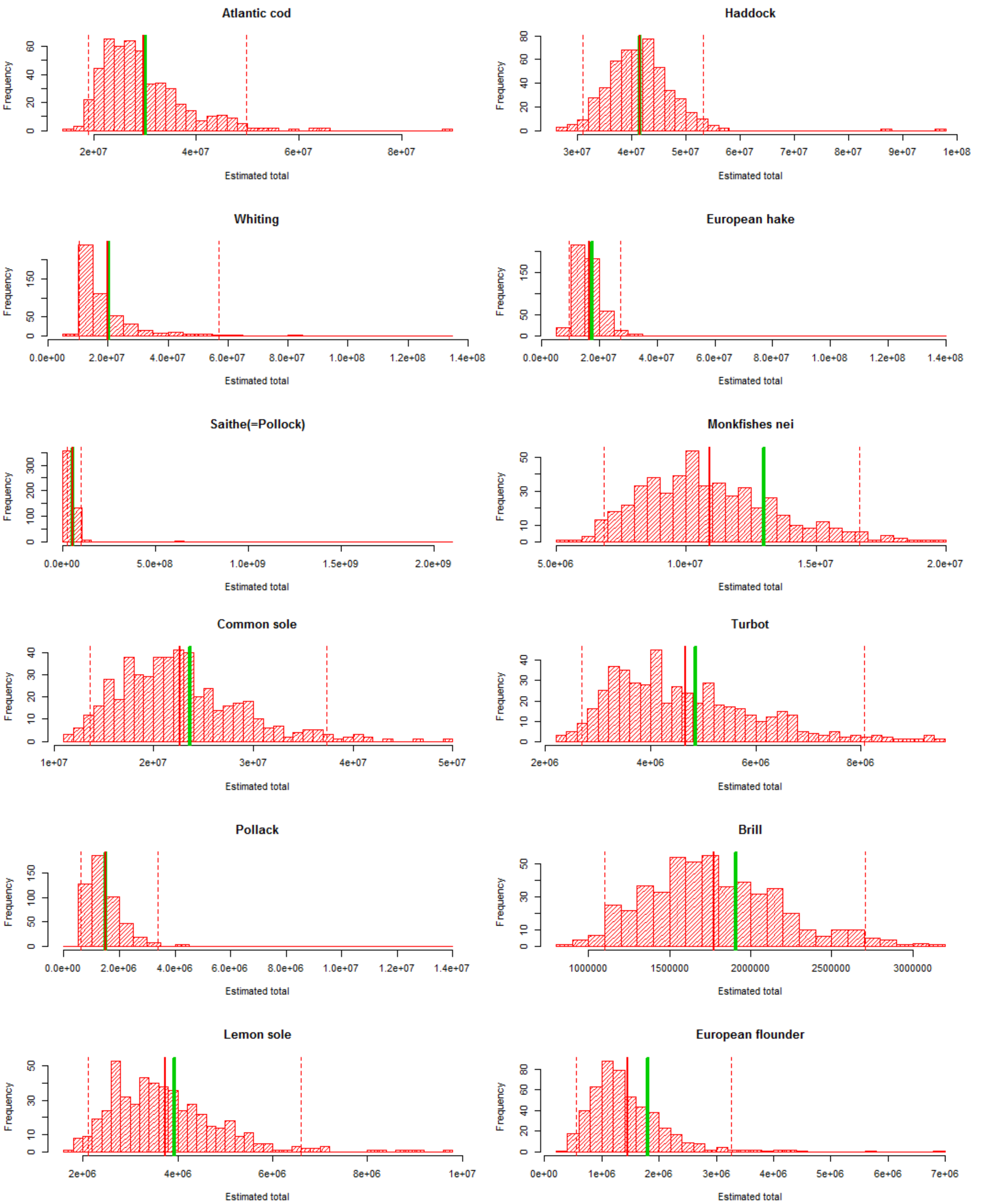


Figure 4: 6 country design: Distribution of 500 estimates of total landed weight for selected species, the mean and 95% confidence intervals of the estimates are marked as red vertical lines, the true value of the species landed weight in green. With the stratification by country and the existing sampling levels, which includes some countries that do no port sampling, there is a noticeable increase in the bias for some species, particularly monkfishes, sole, turbot, brill, flounder.

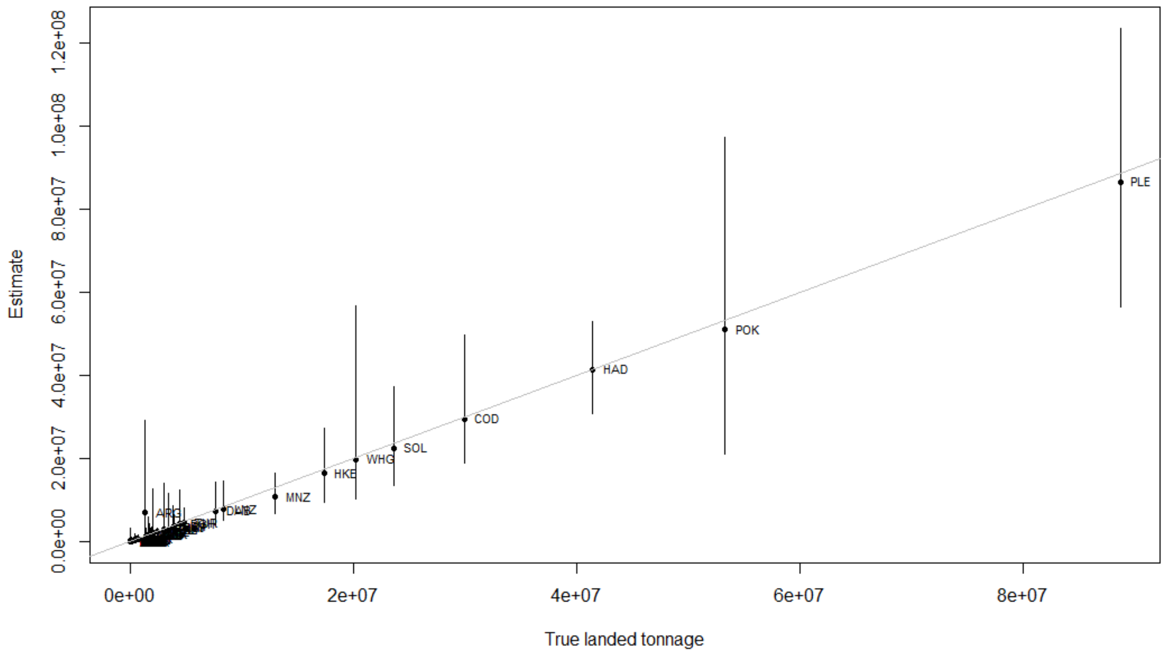


Figure 5: 6 country design: Estimates and 95% confidence intervals of landed tonnages by species. In comparison to the simple random sample when the design is limited to only six countries the variability of saithe shows a considerable decrease while that of whiting has increased.

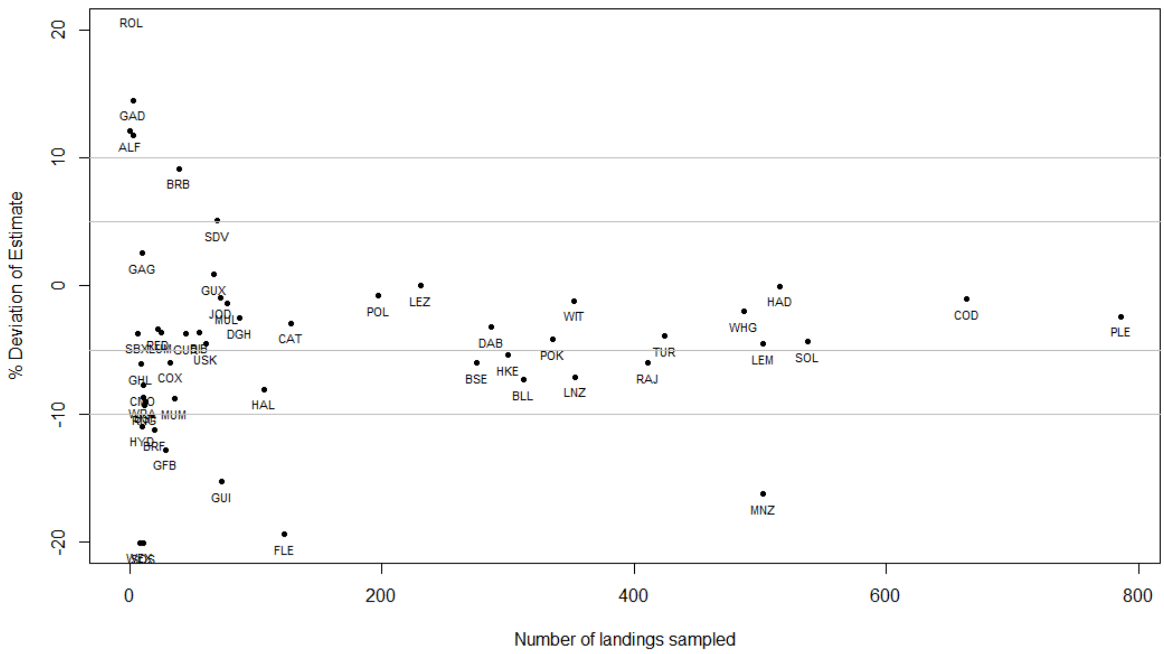


Figure 6:. 6 country design: Percentage deviation of the estimate in relation to the number of samples. In comparison to the simple random sample many of the species show an increase in the deviation of the estimates, monkfish (MNZ) and flounder being particularly notable. This may reflect a systematic bias in the estimates due to poor coverage by some countries.

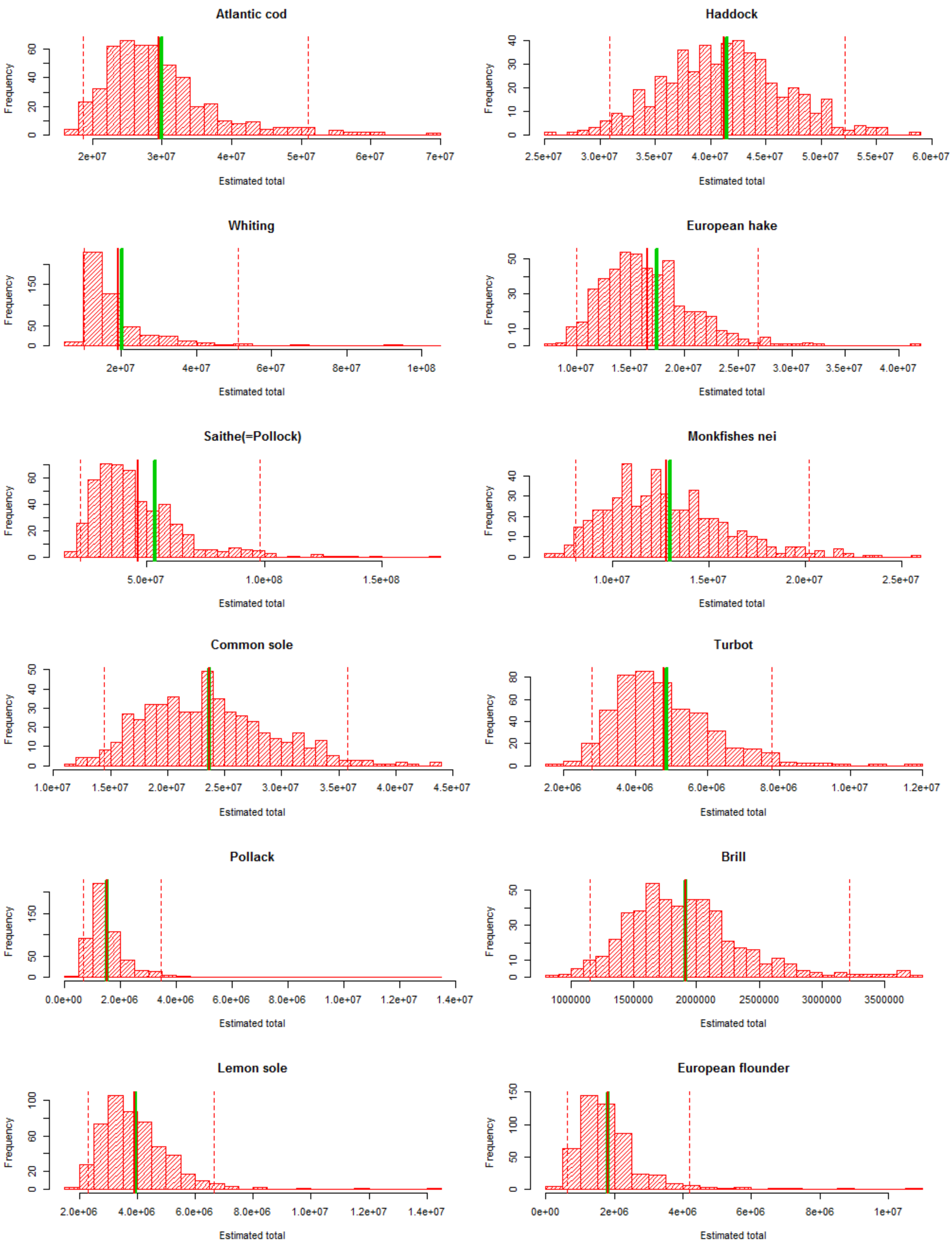


Figure 7: 10 country design: Distribution of 500 estimates of total landed weight for selected species, the mean and 95% confidence intervals of the estimates are marked as red vertical lines, the true value of the species landed weight in green. Increasing the sampling coverage over countries improves the estimates for many of the flatfish species, though increases the deviation for saithe.

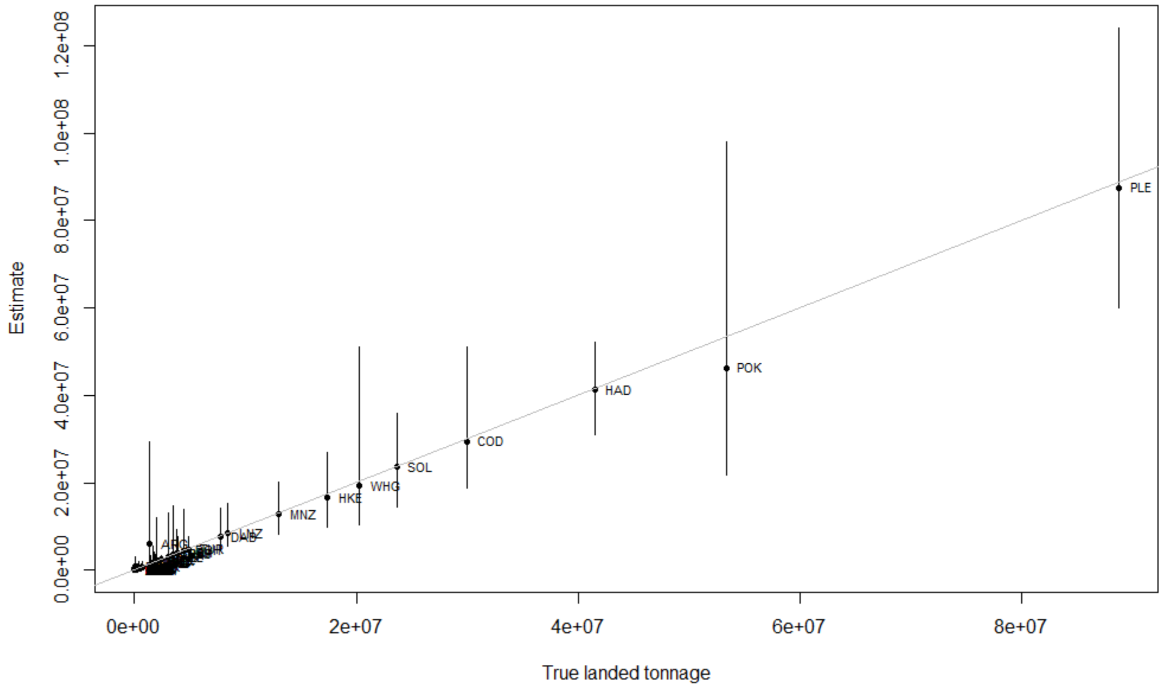


Figure 8: 10 country design: Estimates and 95% confidence intervals of landed tonnages by species. With the ten country design the variation in the saithe estimates is particularly noticeable.

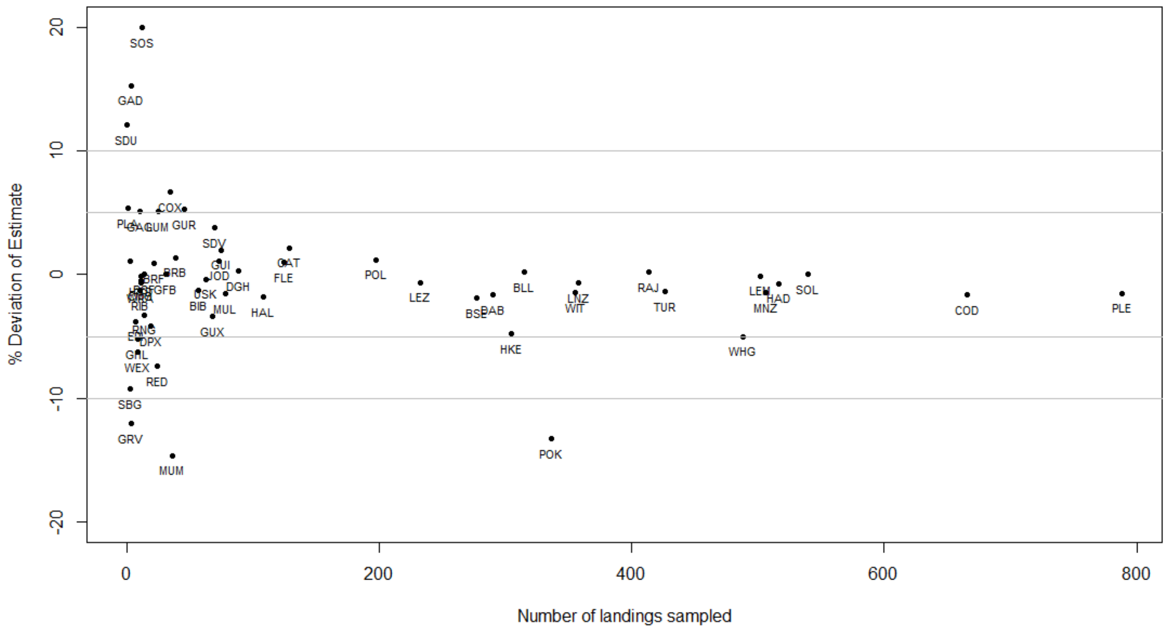


Figure 9: 10 country design: Percentage deviation of the estimate in relation to the number of samples. In comparison to the 6 country design the % deviation for many species has improved, though saithe is a noticeable exception.

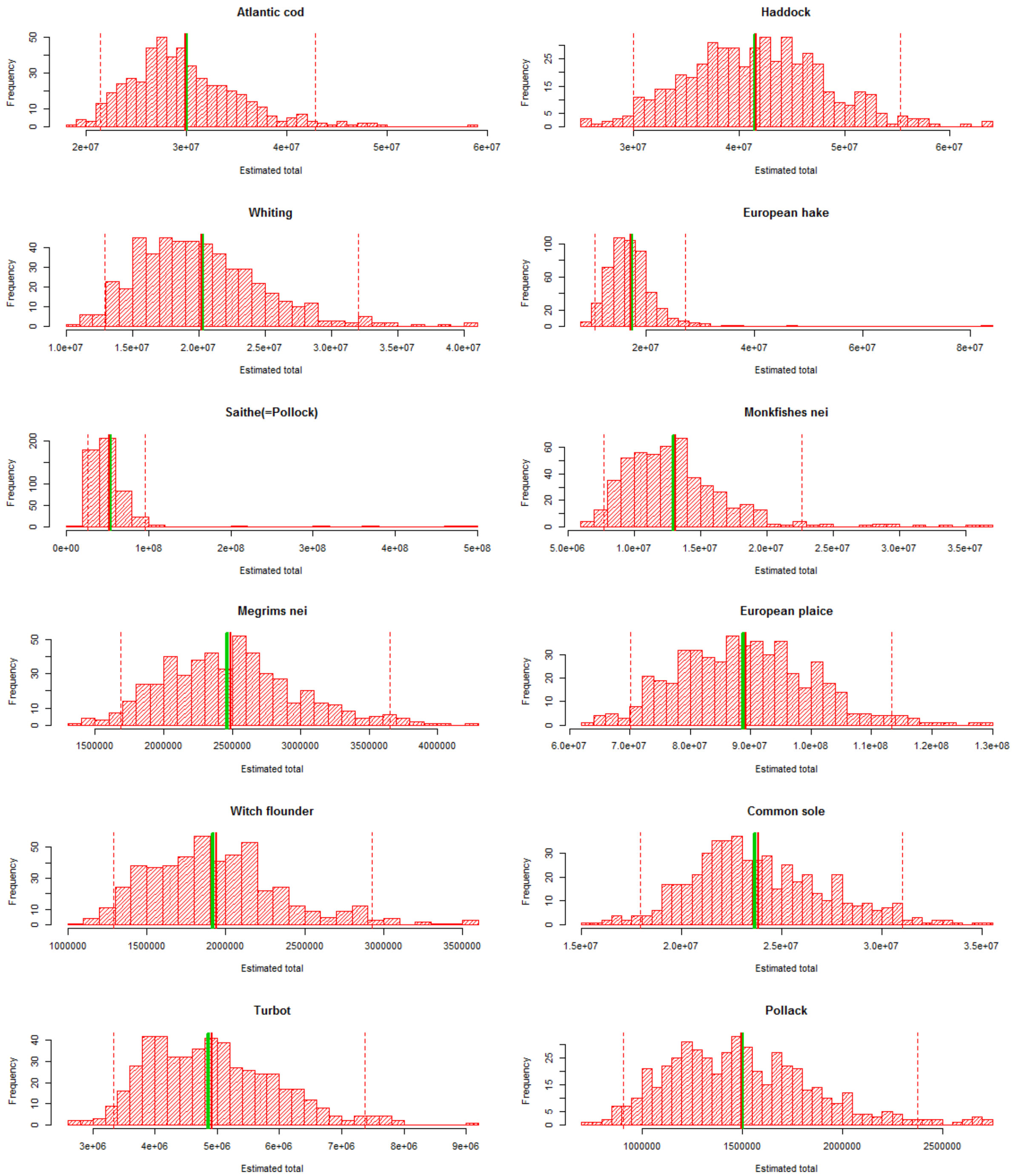


Figure 10: Major and minor ports design: Distribution of 500 estimates of total landed weight for selected species, the mean and 95% confidence intervals of the estimates are marked as red vertical lines, the true value of the species landed weight in green. The stratification by major and minor ports provides unbiased estimates for all the species with significant numbers of samples.

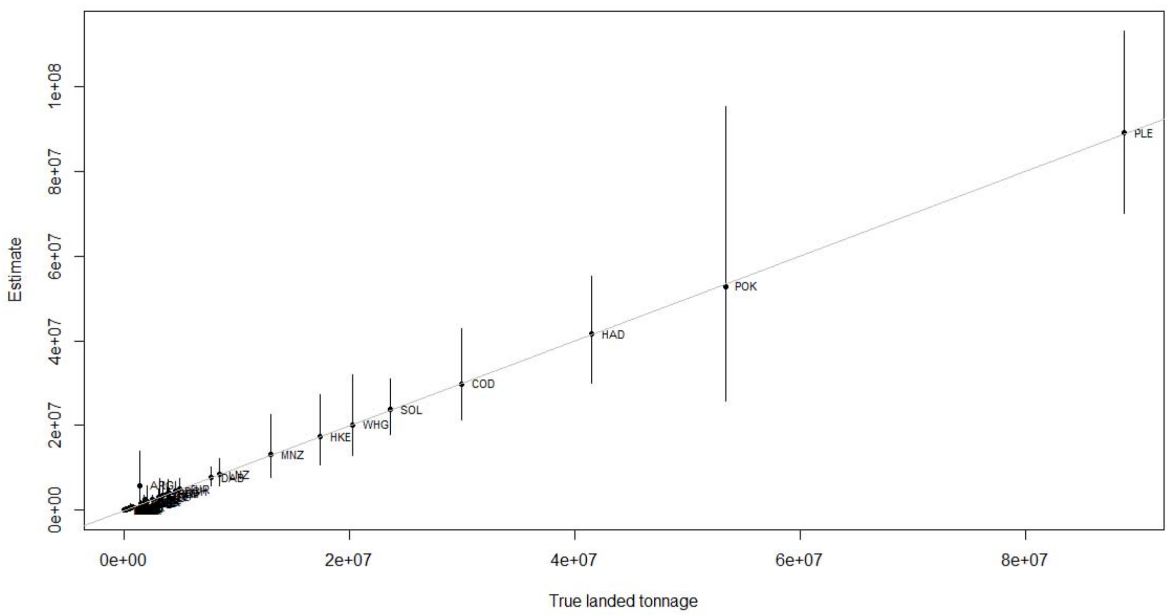


Figure 11. Major and minor ports design: Estimates and 95% confidence intervals of landed tonnages by species. The effort allocation was 800 psu samples from the major port strata and 147 psu from the minor port strata.

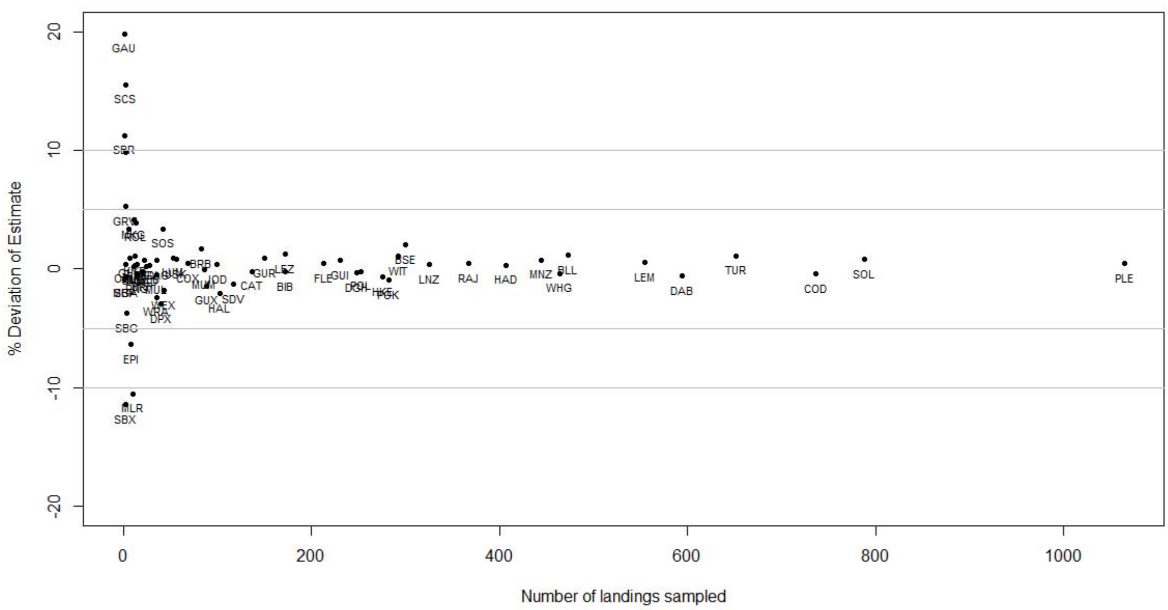


Figure 12. Major and minor ports design: Percentage deviation of the estimate in relation to the number of samples. The major minor port design performs particularly well as most of the species fall within +/- 5% of the true value. The effort allocation was 800 PSU samples from the major port strata and 147 PSU from the minor port strata.

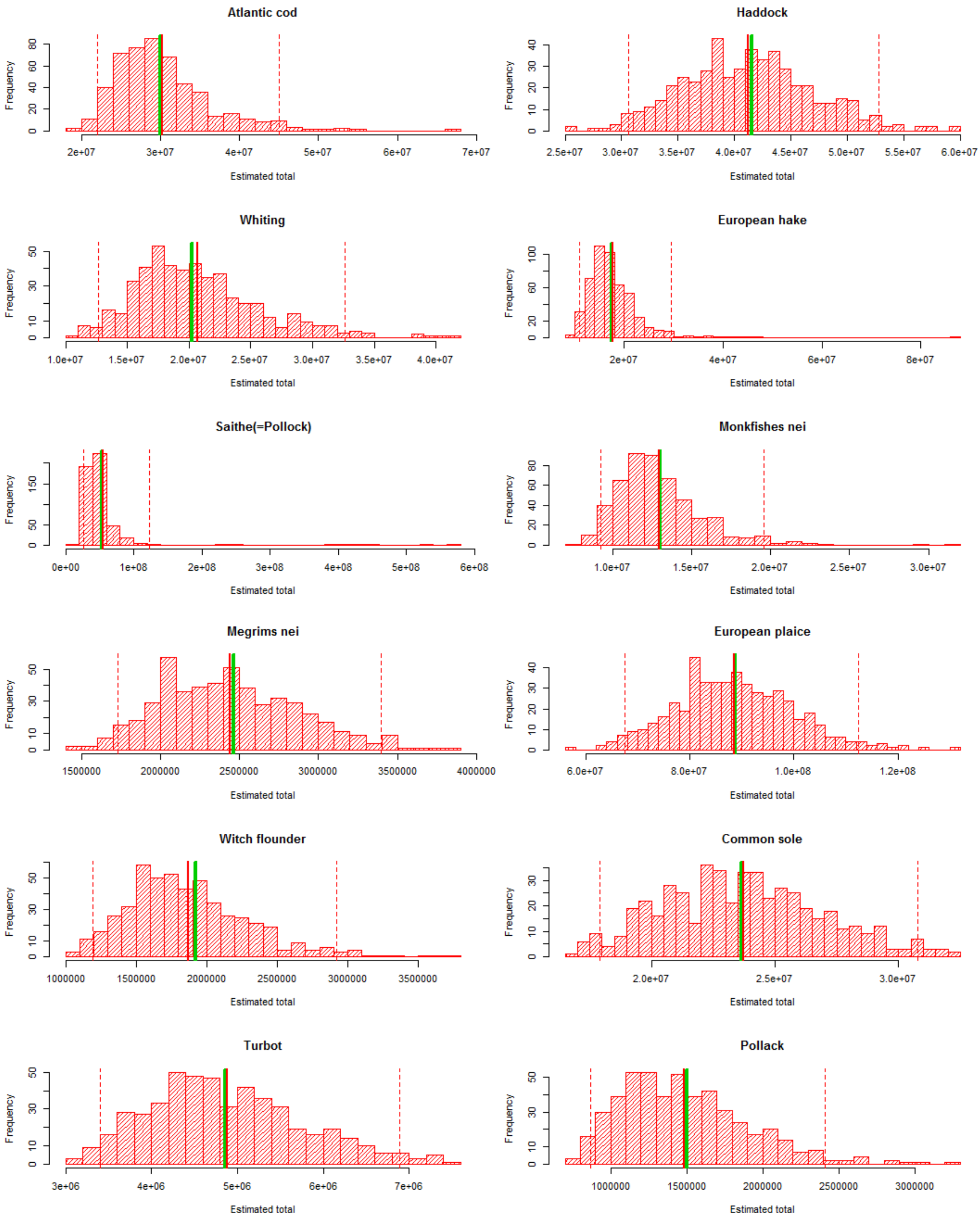


Figure 13. Port and country design: Distribution of 500 estimates of total landed weight for selected species, the mean and 95% confidence intervals of the estimates are marked as red vertical lines, the true value of the species landed weight in green. The stratification by country and major and minor ports provides unbiased estimates for all the species though the RSE for species is not quite as good as that achieved by the major and minor port design.

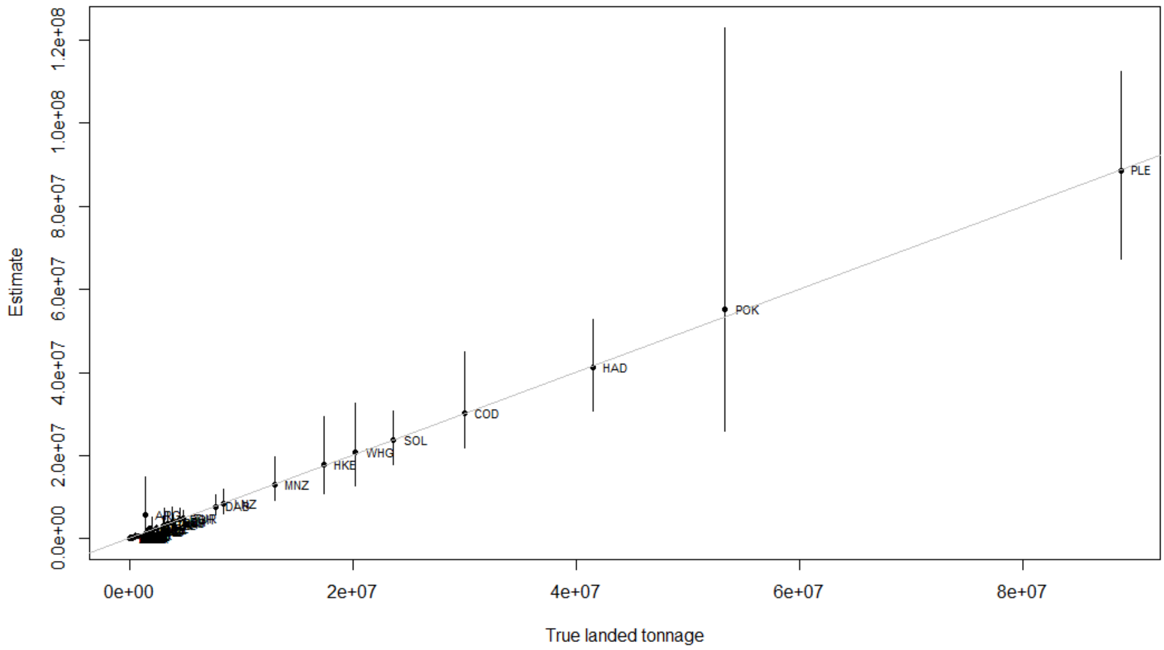


Figure 14. Port and country design: Estimates and 95% confidence intervals of landed tonnages by species. This design generates similar RSE measure across species to the port design but also considerably improved the precision on estimates by country.

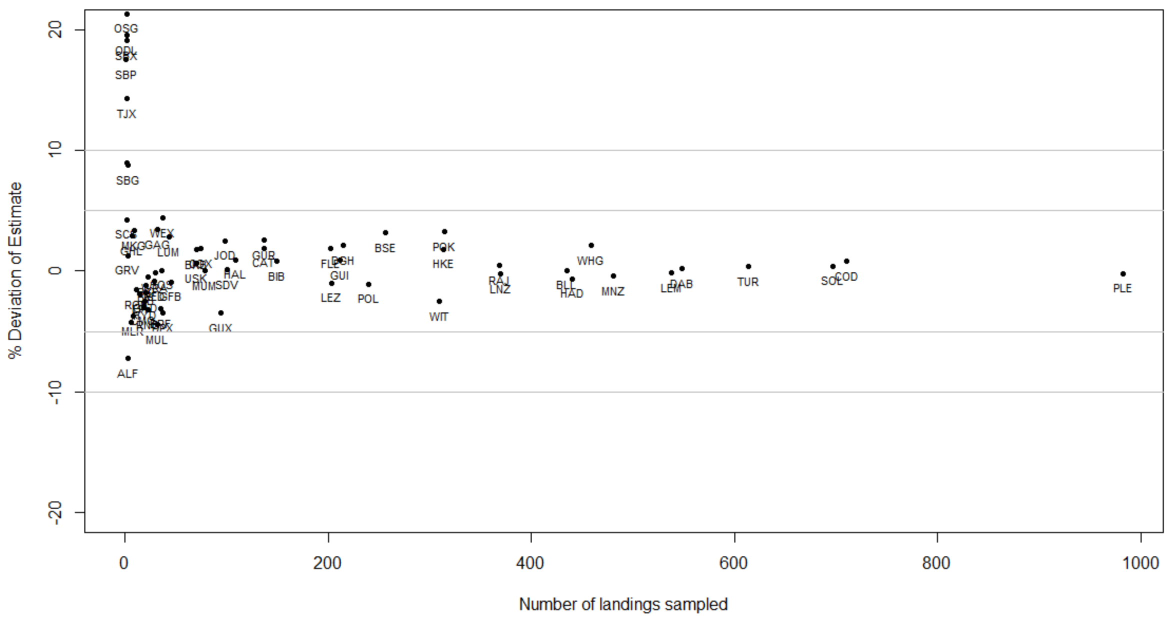


Figure 15. Port and country design: Percentage deviation of the estimate in relation to the number of samples. With the port and country design most of the species fall within +/- 5% of the true value.

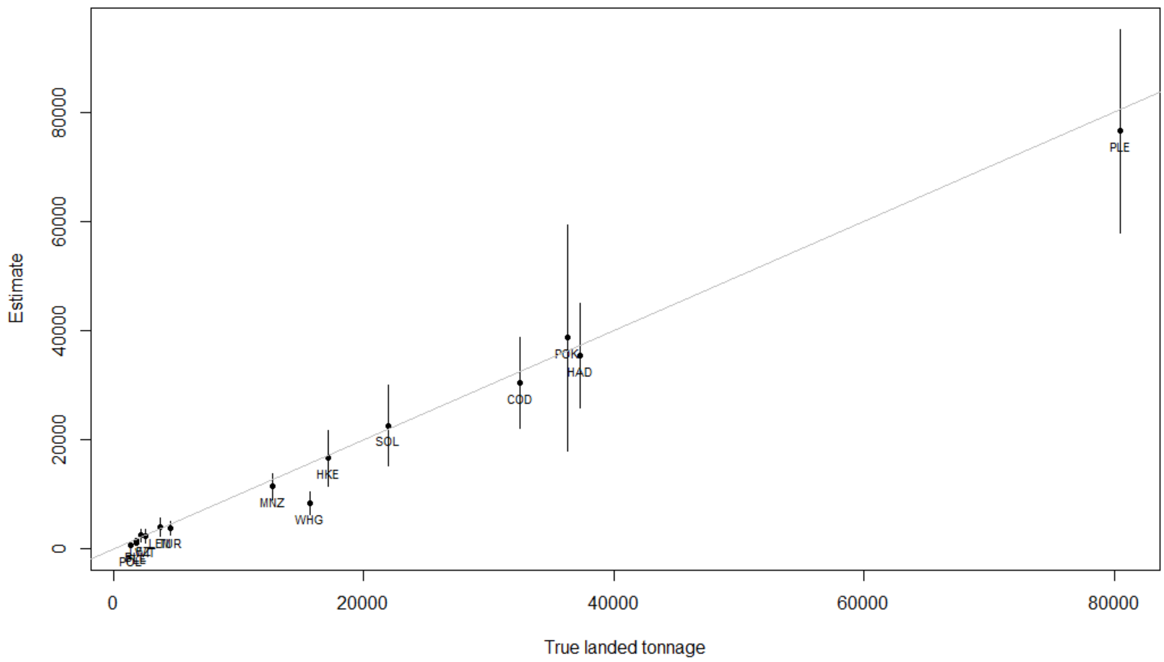


Figure 16: A single realisation of the sampling design used to sample the 2014 data. The point estimates are for the main assessed stocks covered by WKNSSK, the confidence intervals are derived from +/- 1.96 times the standard errors of the estimates. The greatest % deviations are 56% for Brill, and 47% for Witch, where analytical confidence intervals also fail to cover the true value of the landed tonnage. The realised sample was generated from 943 site day sampling units at 142 different landing locations in 11 countries; 1591 individual fishing trips were sampled.

4 At-Sea sampling designs simulations for the North Sea fisheries

fishPi Case Study 2 core team (Alastair Pout, Liz Clarke, Ana Ribeiro Santos, Jon Elson, Patrik Börjesson)

4.1 Introduction

At-sea designs for the collection of regional sampling data have a number of potential advantages over on-shore designs. Firstly, all the catch components, the discarded as well as the landed, are accessible to at-sea observers whereas only the landed fraction is available for on-shore data collection. Secondly, many of the problems and complications of accessing landings at ports, markets and processors do not arise in the at-sea situation. Set against this is the relative cost of at-sea sampling and the limitations on access that may be placed on the observer by the vessel skipper or the working practices of the sampling institutions. Additionally there are the longer time periods involved in the sampling events and the smaller sample size that is generally possible.

Here we assess potential sampling designs for regional at-sea sampling. One aspect of at-sea sampling is the need for discard data collection, and clearly a simulation study based on logbook records of landings cannot be expected to well represent the discarded fraction of the catch. That said, the coverage of fleets, and the relative merits of differing effort allocation schemes can be considered; any at-sea scheme that could not estimate landings well would not be expected to estimate discards well either.

4.2 Methods

The Data

The data used for the simulations was based on collated logbook and sales note data provided by eight countries (Belgium, Germany, Denmark, France, England (inc Wales), Netherlands, Scotland and Sweden) involved in data collection under the DCF data collection regulation. These are the countries operating in the North Sea region. The working data set used for at-sea sampling designs was refined according to area and vessel type (see section 2) and split by year; 2013 and 2014 data being available. The 2013 data consisted of 116,888 voyages (trips), where we define a voyage as a recorded landing by a fishing vessel into a landing location on a particular date; voyages were identifiable to individual vessels and individual vessels make multiple voyages over the course of the year. Individual vessels were given a unique encrypted identification code, and for all vessels the flag country and the length class were available. For the individual voyages, the species landed weight (in kg) was available for all the recorded landed species. Landing location was classified as UN/LOCODE 5 digit location codes and species codes, and dates were harmonised using the criteria and methods set out in section 1. The data set comprises all trips from the North Sea and Northern Shelf areas: ICES divisions IVa, IVb, IVc, VIId and IIIa, VIa and VIb for both 2013 and 2014.

Simulations

The simulations were based on one-stage sampling involving the selection of a vessel and landing date (trip), this is thus the primary sampling unit (PSU). This assumes that although the vessel would be the initial point of contact, the refusal rate would be zero and that an at-sea observer was available to sample the selected trip. For the purposes of the simulations, the data recorded for the selected voyages was the landed weight of all landed species, along with other pertinent details of the vessel and the voyage such as the métier, fishing location etc.

Alternative Designs

The manipulation of the sampling stratification and the distribution of sampling effort using different allocation regimes enabled potential sampling designs to be tested, and evaluated. In total 4 designs with differing sampling strata were evaluated. The overall effort allocation was maintained at ~ 822 trips sampled. For 3 of these designs different methods of allocating sampling effort were compared including using *Neyman's allocation* (Lumley, 2010). Neyman's allocation is the theoretical optimal sampling effort allocation based on the strata with higher variance being allocated a larger sample size. This is dependent on knowing the variance in each of the strata – in practice the data from the previous year was used.

The sampling designs tested were as follows:

1 Simple Random Sampling
No stratification assumed. Trips were selected with simple random sampling without replacement, from all possible trips available. Zero refusal rates were assumed.

2 Eight country sampling
For these simulations, the flag country was used for the stratification. According to the data, there were vessels from 8 flag countries fishing in the North Sea with the stratum totals varying from 1,015 trips from Germany and 37,628 trips from France (Table 1). The distribution of sampling effort for the simulation was based on the present sampling levels for each country. The total number of trips sampled in 2013 was 822, varying from 16 trips sampled by Germany and 354 from France (Table 2). It was assumed that within each flag country each trip was selected randomly with simple random sampling without replacement and the refusal rate was zero.

Table 1: The distribution of unique voyage primarily sampling units by landing country, these formed the stratum totals for the landing country stratification scenario.

Present Number of trips by country (Flag Country)							
BEL	DEU	DNK	FRA	GBR	NLD	SCT	SWE
3151	1015	20187	37628	32227	10433	8729	3038

Table 2: The distribution of sampling effort (PSU) by flag country, these formed the stratum totals for the landing country stratification scenario.

Present Number of trips sampled by country (Flag Country)							
BEL	DEU	DNK	FRA	GBR	NLD	SCT	SWE
33	16	130	354	91	66	60	72

The use of Neyman's sampling effort allocation changed the distribution of the number of trips sampled by each country. Based on the variation in the trip landed weights Neyman's method re allocated more sampling effort to Germany, Scotland and the Netherlands at the expense of Belgium, England and Sweden (Table 3).

Table 3: The distribution of sampling effort (PSU) by flag country, using Neyman's sampling effort allocation.

Neyman's allocation Number of trips sampled by country (Flag Country)							
BEL	DEU	DNK	FRA	GBR	NLD	SCT	SWE
16	182	41	196	41	175	168	3

3 Vessel length: <10m and >=10m
The vessel length stratification by under 10m and over 10m was a two stratum scenario where the flag country was ignored. The distribution of sampling effort for each stratum was based on the landings, maintaining the total number of trips sampled (822). The effort allocated to each stratum was set proportionally to the landed weight, at 804 sampled trips for >=10m and 18 trips for <10m.

When Neyman's sampling allocation was applied (using total landed weight by trip), the results allocated 821 trips for the >=10m stratum and 0.8 (1) trips for the <10m stratum.

4 Country x Vessel length
This scenario linked the flag country with vessel length (>=10m and <10m), which resulted in 15 strata. The present sampling effort (~ 822 sampled trips) by each country was maintained but the distribution of the sampling effort to each of the length classes was proportional to the total landed weight for each (Table 4).

Neyman's allocation (using total landed weight by trip) redistributed the total sampling effort across all strata. This allocation suggested that only the over 10m vessels strata need be sampled, and assigned zero trips to the under 10m strata (Table 4). To ensure the survey package in R worked a nominal 2 trips was allocated to each of the under 10m strata.

Table 4. The effort allocation by country and vessel length.

	Number of trips	
Stratum	**Proportional to landings**	**Neyman's allocation**
BEL-o10m	**33**	**15**
DEU-o10m	**16**	**175**
DEU-u10m	**2**	**(0) 2**
DNK-o10m	**125**	**42**
DNK-u10m	**5**	**(0) 2**
FRA-o10m	**344**	**191**
FRA-u10m	**10**	**(0) 2**
GBR-o10m	**83**	**60**
GBR-u10m	**8**	**(0) 2**
NLD-o10m	**66**	**174**
NLD-u10m	**2**	**(0) 2**
SCT-o10m	**60**	**163**
SCT-u10m	**2**	**(0) 2**
SWE-o10m	**70**	**2**
SWE-u10m	**2**	**(0) 2**

Estimation

The estimation followed the same methodology used for the North Sea on-shore sampling designs (section 2).

4.3 Results

The results for each scenario are shown in Table 5. The relative standard error by species are very similar across the different scenarios, varying from 0.78 for the country design and 0.89 for the regional vessel length. The species with the highest variability in the estimates is POK (saithe) and ARG (Argentines) (Fig. 1). This may be due to the landing patterns and distribution of these species.

In terms of the % deviation in the estimates all the four designs are unbiased (as expected) being all within +/- 1.25% of the true value. The scenarios that performed best were the country x vessel length and country designs, with a % deviation of 0.74% and 0.73%, respectively. The scenario with the highest % deviation was the regional vessel length (-1.02%). Fig. 2 showed that the simulations for each design produced similar outputs. In all four designs the % deviation for each species was low, most species were within +/- 5%. However, the country x vessel length design, provided the least % deviation across species.

The country and vessel length design gives better precision on national estimates than the regional vessel length design, both of which are better than a random design.
The Country x Vessel length (effort allocation proportional to landed weight) appears to offer the best combination of low deviation and precision levels for species and countries. The country stratification offers improvements over the simple random design, but the vessel length design does not improve the estimates compared to the random design.

Table 5. Main outputs of each sampling design

	Design	% Deviation from total	Mean RSE Species	Mean RSE country
1	Random	0.935	0.8207	0.6328
2	Country	0.744	0.7760	0.3116
3	Vessel length (<10, >=10m)	-1.016	0.8930	0.3039
4	Country x Vessel length	-0.021	0.8344	0.2519

When we applied Neyman's sampling effort allocation to the Vessel length and Country x Vessel length scenarios, the results showed an improvement of the overall percentage deviation for the vessel length design, but an increase in the overall percentage deviation in Country x Vessel length design (Table 6). Neyman's allocation for both scenarios assigned zero trips to be sampled on the under 10m strata. Neyman's sampling effort allocation altered the current sampling effort distribution. According to Neyman, Germany, Scotland and the Netherlands should increase the sampling effort for the over 10m vessels, while France, Denmark and England should decrease their sampling effort. The most dramatic change was the increase of sampled trips from Germany, from 16 sampled trips to 175 trips (Table 3).

Table 6. Main outputs of each sampling design using Neyman's allocation

	Scenario	% Deviation from total	mean RSE species	mean RSE country
2	Country	-0.439	0.8158	0.5061
3	Vessel length	0.731	1.0005	0.3390
4	Country x Vessel length	-1.258	0.9485	0.45152

Table 7 shows the individual estimations for the species with the highest commercial interest in the North Sea for the Country x Vessel length scenario using the 2013 data and effort proportional to landings. For all these species the mean estimations are close to the true value and presented low % deviation, varying between 0.064% for plaice and -3.3% for saithe.

Table 7. Species landed weight “true” values from the known data set, and the estimated values, standard error, % deviation, relative standard error and 95% confidence interval from 500 simulations using the port and country sampling design.

Species		*True value*	*Estimate*	*SE*	*% Deviation*	*RSE*	*95% CI*	
							Lower	**Upper**
Seabass	*Dicentrarchus spp*	3,099	3,103	1,063	0.101	0.343	1,486	5,489
Cod	*Gadus morhua*	29,976	30,071	4,782	0.316	0.159	21,109	40,374
Witch	*Glyptocephalus cynoglossus*	1,921	1,904	456	-0.881	0.239	1,175	2,958
Megrims nei	*Lepidorhombus spp*	2,463	2,491	813	1.137	0.326	1,147	4,418
Dab	*Limanda limanda*	7,713	7,682	991	-0.402	0.129	5,923	9,819
Monkfish nei	*Lophius spp*	13,013	12,890	6,934	-0.944	0.538	5,392	32,686
Haddock	*Melanogrammus aeglefinus*	41,501	41,489	8,119	-0.028	0.196	27,845	58,786
Whiting	*Merlangius merlangus*	20,267	20,340	3,440	0.360	0.169	14,326	27,651
Hake	*Merluccius merluccius*	17,410	17,146	5,780	-1.516	0.337	8,138	30,086
Lemon sole	*Microstomus kitt*	3,926	3,906	660	-0.515	0.169	2,707	5,231
Flounder	*Platichthys flesus*	1,806	1,831	573	1.375	0.313	980	3,161
Plaice	*Pleuronectes platessa*	88,777	88,720	10,398	-0.064	0.117	69,945	108,817
Pollack	*Pollachius pollachius*	1,502	1,498	460	-0.233	0.307	773	2,567
Saithe	*Pollachius virens*	53,363	51,563	26,956	-3.374	0.523	23,253	138,538
Turbot	*Psetta maxima*	4,861	4,859	768	-0.026	0.158	3,527	6,408
Brill	*Scophthalmus rhombus*	1,913	1,917	261	0.215	0.136	1,417	2,429
Sole	*Solea solea*	23,631	23,490	2,681	-0.598	0.114	18,429	29,604

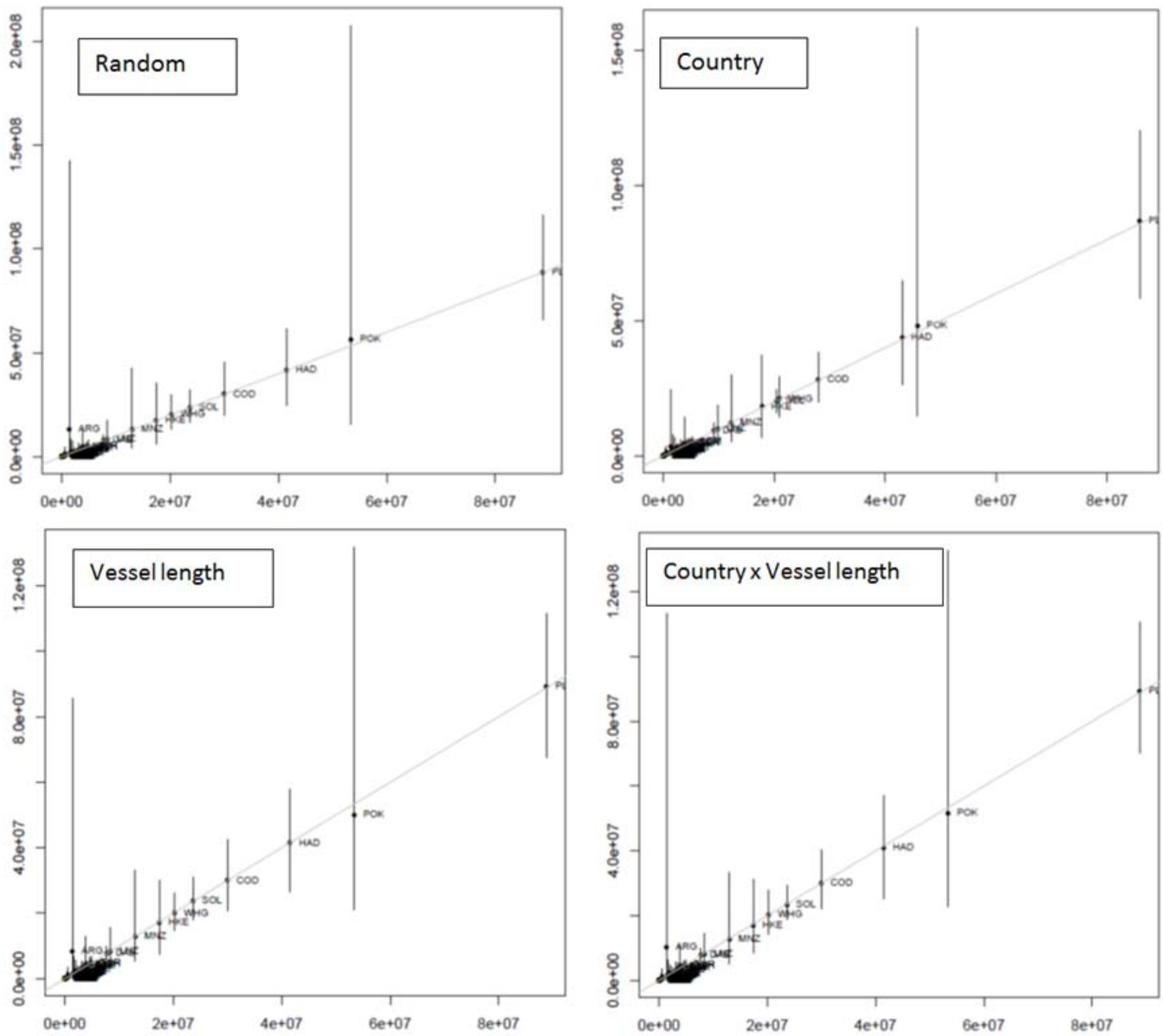


Figure 1. Comparison of the four at-sea sampling designs: scatter plots of the estimated total landed weight by species (y axis), against the true landed weight (x axis), confidence intervals are generated from the 95% percentiles of 500 simulations.

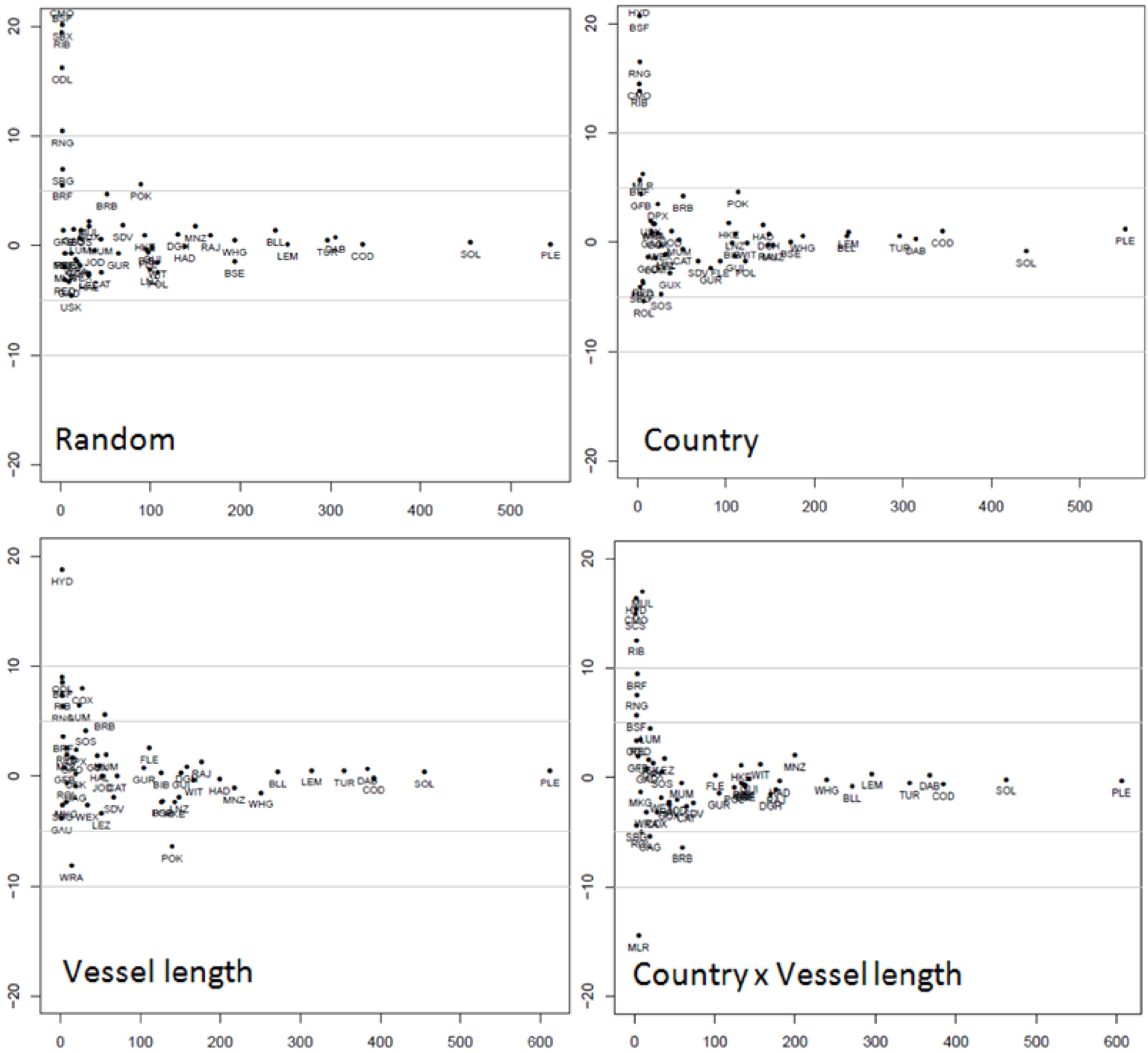


Figure 2. Comparison of the four at-sea sampling designs: scatter plots of the % deviation (y axis) against the sample size(x axis), the greater the sample size the less the % deviation in estimates with the majority of the species falling within +/- 5% of the true value.

The optimal design, country x vessel length (sampling effort proportional to landings), was applied to 2014 data to generate a single estimate for the main commercial species in the North Sea. The % deviation in the estimates ranging from -44% for saithe to 33% for haddock. Most of the species considered had % deviation values in excess of +/- 10% (Table 8).

Table 8: Species landed weight "true" values from the known data set, and the estimated values, standard error, % deviation, relative standard error and estimated 95% confidence interval for a single simulation using the country x vessel length sampling design.

Species		True value	Estimate	SE	% Deviation	RSE
Cod	*Gadus morhua*	32519	37048	6074	13.929	0.164
Witch	*Glyptocephalus cynoglossus*	2547	1978	399	-22.317	0.202
Megrims nei	*Lepidorhombus spp*	2192	2391	687	9.122	0.287
Dab	*Limanda limanda*	5506	4872	455	-11.516	0.093
Monkfish nei	*Lophius spp*	12689	11976	2607	-5.614	0.218
Haddock	*Melanogrammus aeglefinus*	37283	49527	9506	32.841	0.192
Whiting	*Merlangius merlangus*	15704	17951	2516	14.303	0.140
Hake	*Merluccius merluccius*	17125	11370	3511	-33.609	0.309
Lemon sole	*Microstomus kitt*	3688	3225	410	-12.555	0.127
Flounder	*Platichthys flesus*	1828	2021	559	10.536	0.276
Plaice	*Pleuronectes platessa*	80507	94501	10600	17.382	0.112
Pollack	*Pollachius pollachius*	1378	1076	388	-21.907	0.361
Saithe	*Pollachius virens*	36312	20226	7241	-44.299	0.358
Turbot	*Psetta maxima*	4573	5053	1098	10.482	0.217
Brill	*Scophthalmus rhombus*	1793	1943	270	8.397	0.139
Sole	*Solea solea*	22000	20473	2621	-6.943	0.128

4.4 Discussion

At-sea observer trips are used to provide estimates of catch and discards by age and species composition. The population data available for this study is landed weight which is used as a proxy for the biological data required. Any assessment of the suitability of the designs to estimate the discarded fraction of the catch has to assume that the discards are directly proportional to the landings which may not be the case for all bycatch species. However, as previously mentioned, it is likely that a design which does not result in good estimates of landings well will not result in good estimates of discards.

Neyman's sampling allocation provides the stratum with the most variable elements with proportionally the greatest sampling effort in order to improve the precision of the overall estimate, here, total landed weight, and will therefore not necessarily improve precision for individual design strata, such as country, or other parameters, such as total landed weight by species, as shown by the results presented here. The criteria used here was the variance in the total landed weight by trip for 2013 – the same data used in the simulations, so in retrospect it would have provided a more rigorous test of the designs had they been tested on the 2014 data.

Country and vessel length as a stratification provides, by default, a spatial element to the stratifications; coastlines, landing and departure ports, areas fished, fishing practices and catch and species composition are all going to relate to the flag country and the size of the vessel. That said it is clear that some flag fleets are more mobile, more nomadic than others, and that this tendency increases with the size of the vessel. Vessel length will have a strong influence on catching capacity, trip duration trip frequency, and fishing areas. Additionally under 10m and over 10m vessels are often selected as stratum because of the different reporting procedures and management measures.

Trip length will play an important part in the logistics of any design implementation as it has a direct influence on costs; there is a clear difference between the cost of observing 822 day trips and 822 week long trips.

Further studies of at-sea sampling designs could focus on distinct fleet segments perhaps categorized by the vessel type variable defined in the data set used here. It would also be instructive to explore discard and by-catch sampling by providing estimates and sample size by métier. Critical decisions as to the suitability of sampling designs require the engagement of end users of the data, and in the at-sea situation there are a number of potential end users with differing data needs.

References

Pastoors, M., Quirijns, F. (2014). Discard Atlas of North Sea fisheries. Scheveningen group. 84pp.

Rogers, S. and Stocks R (2001). North Sea fish and Fisheries. Technical Report TR_003. 72pp.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Lumley, T. (2014) Survey: analysis of complex survey samples. R package version 3.30.

Kish, L. (1965) Survey Sampling. John Wiley & Sons, New York.

Horvitz, D. G.; Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *Journal of the American Statistical Association*, 47, 663–685.

ANNEX 13 WP2 CS3 North Sea flatfish fisheries sampling designs

The fishPi project
Regional co-operation in fisheries data collection

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2.3 – Case Study fisheries

Deliverable CS3 Southern North Sea flatfish fisheries sampling designs

Lead: IMARES, Netherlands.

Core Team: Chun Chen, Sieto Verver, Edwin van Helmond (IMARES, Netherlands)
Ana Ribeiro Santos, Jon Elson (CEFAS, UK England)
Els Torreele, Sofie Nimmegeers (ILVO, Belgium)
Julia Wischnewski (TI-SF, Germany)

Data providers: IMARES (Netherlands), CEFAS (United Kingdom), ILVO (Belgium), TI-SF (Germany), MSS (United Kingdom), DTU-Aqua (Denmark), IFREMER (France), SLU (Sweden).

Table of Contents

1 Introduction

Flatfish fisheries in the North Sea area (includes North Sea, Skagerrak and Eastern Channel) is primarily focused on the target species sole (*Solea solea*) and plaice (*Pleuronectes platessa*) in the southern areas of the North Sea. Turbot (*Scophthalmus maximus*) and brill (*Scophthalmus rhombus*) are valuable by-catch products, other frequently caught flatfish species are commonly of low value and regularly discarded, e.g. dab (*Limanda limanda*), flounder (*Pleuronectes flesus*). Fleets involved originate from different countries, the Netherland, Belgium, England, Scotland, France, Denmark, Sweden and Germany. Gears generally used are beam trawlers or otter trawlers, also, a small scale fishery using passive gear (gillnets and trammel nets) and a small group of vessels using demersal (Scottish seine) seines to target plaice.

Information collection on biological parameters of catch is legislated via the Data Collection Framework (DCF) and the Data Collection Multiannual Programme (DCMAP). This is done through national on-shore and at-sea monitoring programmes. Trained personnel collect the biomass, length, age and species composition of the catch, landings and discards, from their most important commercial fisheries, with the main aim to feed these data into stock assessments. All data are stored and administrated by respective national authorities. Synthesizing data from as many different fisheries, regions and countries as possible is required to facilitate stock assessments and, also, European-wide management approaches. So far, such synthesis were hampered by (i) the diversity of procedures in collecting and processing data, (ii) the disparate intensities of sampling compared with the total fishing effort across countries, (iii) the lack of common data exchange format and storage facility, and (iv) national regulations which precluded sharing of detailed commercial catch data (Uhlmann, 2014; ICES, 2011).

For more than fifty years stock assessments are conducted on North Sea stocks of sole and plaice (ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak). Since 2013, a stock assessment is in place for North Sea turbot. Only since 2004, estimates of discard numbers are included in the stock assessment for plaice. For all other flatfish species discard information is collected but are not included in stock assessments, mainly driven by reasons as summarized above, e.g. diversity of procedures, disparate intensities of sampling, the lack of common data exchange formats.

The current situation of data collection in North Sea flatfish fishery will form the basis for a case study in analysing and evaluating different sampling designs in a regional framework. The output of the case study will lead, eventually, to improved data collection through strengthening regional cooperation in fisheries data collection.

2 Flat fish fisheries in the North Sea area

2.1 Physical and biological environment of the North Sea area IIIa, IV and VIId

Based on the differences in bathymetry, topographical features and hydrography, a number of different habitats and regions can be identified in the North Sea (Figures 1a and 1b).

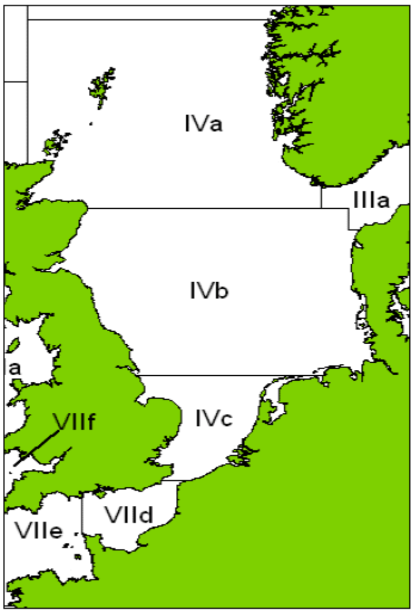


Figure 1a. North Sea overview (Discard Atlas North Sea Fisheries, 2013) **Figure 1b**. North Sea divided in ICES areas.

The sea bottom of the North Sea consists of a variety of sand and mud sediments with small patches of gravel and pebbles. Fine sand predominate over wide areas giving way to soft silt clay muds in some of the deeper areas. Areas of gravel and pebbles are most prevalent to the Southeast of England and off the Danish coast. The hydrography of the area is strongly influenced by inflow of Atlantic water to the North of Scotland and also water from the channel. Together with water draining into the North Sea from numerous large rivers, the overall nutrient input generates a productive environment supporting a number of commercially-important fish and shellfish stocks.

2.2 General fisheries description

A variety of fisheries and fleets, with a variety of gear types is used in the North Sea region. The extent to which different countries participate in the various fisheries depends to a large extent on national quotas available to them.

Based on the International Standard Statistical Classification for Aquatic Animals and Plants (ISSCAAP), 23 different flatfish species were landed in the North Sea area (ICES areas III, IV and VIId) in 2014. In total landings of these species was 128633 tonnes. The main flat fish species in diminishing order are European plaice, common sole, common dab, Greenland halibut, lemon sole and witch. Important landing ports are Harlingen, Den Helder, IJmuiden, in the Netherlands, and Thyborøn in Denmark (see also Table 3). The recorded métiers show a predominance of the beam trawlers in this fishery, accounting for 64006 tonnes (50%) of the landings, other important métiers are otter trawlers and gillnetters. The main fleets involved are the Dutch flagged vessels with vessel length over 40 meters, Dutch flagged vessels with vessel length between 24 and 40 meters, English flagged vessels over 40 meters and Belgian flagged vessels between 24 and 40 meters. The majority of the landings originate from the southern North Sea.

2.3 Fisheries using beam trawls, otter trawls, seines and fixed gear

Beam trawlers with mesh size above 120 mm: The larger meshed beam-trawl gear is principally used in the plaice fishery of the Central and Eastern North Sea. Cod is also fished with this fishery. Denmark, Belgium and England are the main operators of this fishery.

Beam trawlers with mesh size between 80 and 120 mm: This gear is accounting for around 40% of all fishing effort in the North Sea. It is mainly used in a fishery located in most Southerly parts of the North Sea and into the Channel. This fisheries targets a mix of sole, plaice and other flatfish, and is operated principally by the Netherlands, Belgium and Germany.

Otter trawlers with mesh size between 70 and 100mm: The use of this gears are more widespread than the beam gear and associated mainly with three kind of fisheries. i) A fishery targeting Norway lobster (*Nephrops*). These crustacean live on areas of soft clay muds which are distributed patchily throughout the North Sea and Skagerrak. Bycatch limits for fish species apply in the smaller meshed (80-89 mm) *Nephrops* fishery. The bycatch limits do not create undue problems in inshore areas where fish abundance is low. In more northerly offshore areas where fish are more abundant, adhering to the bycatch limits is more challenging. Flatfish species are a regular bycatch in this fishery. ii) A mixed fishery taking place in the more southerly parts of the North Sea and centred on the Eastern Channel in which whiting and non-quota species are important stocks fished. This is predominantly a French fishery. iii) A fishery with 90-99 mm mesh, a mixed demersal fishery more in the centre of the Skagerrak. This is predominantly a Danish and Swedish fishery. In the Skagerrak, also a fishery targeting *Nephrops*, with sorting grid (70-89 mm mesh size), is operated by Swedish vessels.

Beam trawlers with mesh size between 16 and 32 mm: This fishery targets mainly brown shrimp (*Crangon crangon*). Two distinct areas can be identified: in the South, and off the German, Dutch and Belgian coasts. Flatfish species are a regular bycatch in this fishery.

Otter trawlers with mesh size above 100 mm: This gear is predominantly used in the more northern parts of the North Sea. At least three different fisheries operate within this gear category. i) A mixed demersal fishery targeting cod and associated species (mainly haddock and whiting in the Western and Northern North Sea, mainly plaice in the South-eastern North Sea) with trawls and seines nets. The main countries involved are Scotland, Denmark and Germany. ii) A mixed fishery that is characterised by a greater preponderance of 'groundfish' species targeting in particular anglerfish and megrim. This fishery is particularly important in Scotland. iii) A fishery for saithe, with French, German and Norwegian vessels as the main players.

Demersal seiners, also called fly-shooters or Scottish seine with mesh size between 80 and 135 mm: A fishery that targets a mix of demersal species, including the important flat fish species. Operates from the southern parts of the North Sea and the Eastern Channel up to northern central part of the North Sea. A small fleet of a dozen vessels operates form the Netherlands.

Fisheries using fixed gear: A number of fixed gears are used in the North Sea, the most important being gill nets and trammel nets. The main gillnet activity is from a Danish fishery targeted mainly at cod and plaice. The importance of anglerfish in this fishery has increased in recent years and activity directed at this species has increased by mainly Scottish vessels. Trammel net fisheries are operated by a number of countries and are particularly important in more coastal waters for example off the English North Sea and Channel coasts for sole. Denmark has trammel net fishery targeting plaice and cod. Some small scale fisheries using longlines catch cod, hake and ling.

Fisheries using other gears: Most countries also have inshore fisheries conducted by under 10m vessels using a variety of gears (including pots, dredges etc.) for a variety of fish and shellfish species.

2.4 Management and Assessment

TAC & quota by country

Total allowable catches (TACs) or fishing opportunities, are catch limits (expressed in kg, tonnes or numbers) that are set for most commercial fish stocks. The Commission prepares the proposals, based on scientific advice on the stock status from advisory bodies such as ICES and the Scientific, Technical and Economic Committee for Fisheries (STECF). Some multi-annual plans contain rules for the setting of the TACs. TACs are set annually for most stocks (every two years for deep-sea stocks) by the Council of fisheries ministers. For stocks that are shared and jointly managed with non-EU countries, the TACs are agreed with those (groups of) non-EU countries.

TACs are shared between EU countries in the form of national quotas. For each stock a different allocation percentage per EU country is applied for the sharing out of the quotas. This fixed percentage is known as the relative stability key. EU countries can exchange quotas with other EU countries.

EU countries have to use transparent and objective criteria when they distribute the national quota among their fishermen. They are responsible for ensuring that the quotas are not overfished. When all the available quota of a species is fished, the EU country has to close the fishery (http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R0104&from=EN).

TAC and quota for 2015 are fixed by Council Regulations (EU) No 1221/2014 of 10 November 2014, No 1367/2014 of 15 December 2014, No 2015/104 of 19 January 2015, and No 2015/106 of 19 January 2015. Changes may be made during 2015. In table 1; the TAC and quota is given for the four commercially most important species: Common sole (*Solea solea*), Plaice (*Pleuronectes platessa*), and Turbot (*Psetta maxima*) and Brill (*Scopthalmus rhombus*), together under one TAC.

An overview of all TAC and Quota for all species fished in the North Sea Region is found on the website of EU, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R0104&from=EN

Table 1: TAC and Quota for the four species in CS3, by country. TAC and quota are in 1000 kg.

species	Zone	TAC	BE	DK	GE	FR	NL	SE	UK
Solea solea	III	205		172	10		17	6	
	IIa,IV	11900	991	453	793	198	8945		510
	VIId	3483	938			1875			670
Pleuronectes platessa	Skagerrak	10056	60	7830	40		1506	419	
	Kattegat	2626		2337	26			263	
	IV	128376	7365	23938	6905	1381	46035		34066
	VIIde	4787	783			2611			1393
Psetta maxima & *Scopthalmus rhombus*	III								
	IIa,IV	4642	340	727	186	88	2579	5	717
	VIId								

Stock assessments

In total 240 different stock assessments are conducted by ICES on a regular basis. This includes 9 stock assessments on flatfish species in the North Sea area (Table 2). These assessments are conducted by 3 different expert working groups. All the stocks except one, turbot in division IIIa (Skagerrak), are considered to be data rich stocks (target stock category 1). For turbot in division IIIa a survey trends based assessment is used.

Table 2. Description of the stock assessments on plaice, sole, turbot and brill in the North Sea area (III, IV and IVd), includes information on the responsible expert working groups, target stock category and which biological parameters are used for the assessment.

Name	ICES stock code	Workgroup	Assessment Type	Assessment Model	Target stock category	Length freq.	ALKs	Growth information	Weight	Sex ratio	Maturity	Fecundity
Brill in Subarea IV and Divisions IIIa and VIId,e	bll-nsea	wgnew	Survey trends based assessment		1	Yes, annually	Yes, annually	-	Yes, annually	No	Yes, annually	No
Plaice in Division VIId (Eastern Channel)	ple-eche	wgnssk	Trends assessment	XSA	1	Yes, quarterly	Yes, quarterly	not specific	Yes, quarterly	No		No
Plaice Subarea IV (North Sea)	ple-nsea	wgnssk	Analytic	XSA	1	Yes, quarterly	Yes, quarterly		Yes, quarterly	No	Not used currently but will be needed in shot-medium term	No
Plaice in Subdivision 20 (Skagerrak)	ple-skag	wgnssk	Trends survey	-	1	Yes, quarterly	Yes, quarterly	nothing more than from ALK needed	Yes, quarterly	No	Not used and no need to be collected, for the time being	No
Sole in Division VIId (Eastern Channel)	sol-eche	wgnssk	Analytic	XSA	1	Yes, quarterly	Yes, quarterly	-	Yes, quarterly	Yes, quarterly	Other	No
Sole in Division IIIa and Subdivisions 22-24 (Skagerrak, Kattegat, and the Belts)	sol-kask	wgbfas	Analytic	SAM	1	Yes, quarterly	Yes, quarterly	-	Yes, quarterly	Not used currently but will be needed in shot-medium term	Yes, quarterly	Not used and no need to be collected, for the time being
Sole in Subarea IV (North Sea)	sol-nsea	wgnssk	Analytic	XSA	1	Yes, quarterly	Yes, quarterly	-	Yes, quarterly	Not used and no need to be collected, for the time being	Not used currently but will be needed in shot-medium term	No
Turbot in Division IIIa	tur-kask	wgnew/wgnssk	Survey trends based assessment	-	3	Yes, annually	Not used and no need to be collected for the time being		Yes, annually	Not used and no need to be collected, for the time being	Yes, annually	no
Turbot in Subarea IV	tur-nsea	wgnssk	Analytic	SCA (poss trends only)	1-2	Yes, quarterly	Yes, quarterly	available from benchmark	Yes, annually	Not used and no need to be collected, for the time being	Not used and no need to be collected, for the time being	No

2.5 Fisheries description based on fishPi data

Landings by species

Average total annual landings between 2013 and 12014 in weight (tonnes) per species (taxon) are given in Figure 2. Plaice, *Pleuronectes platessa*, is by far the most abundant species in the landings for the North Sea area.

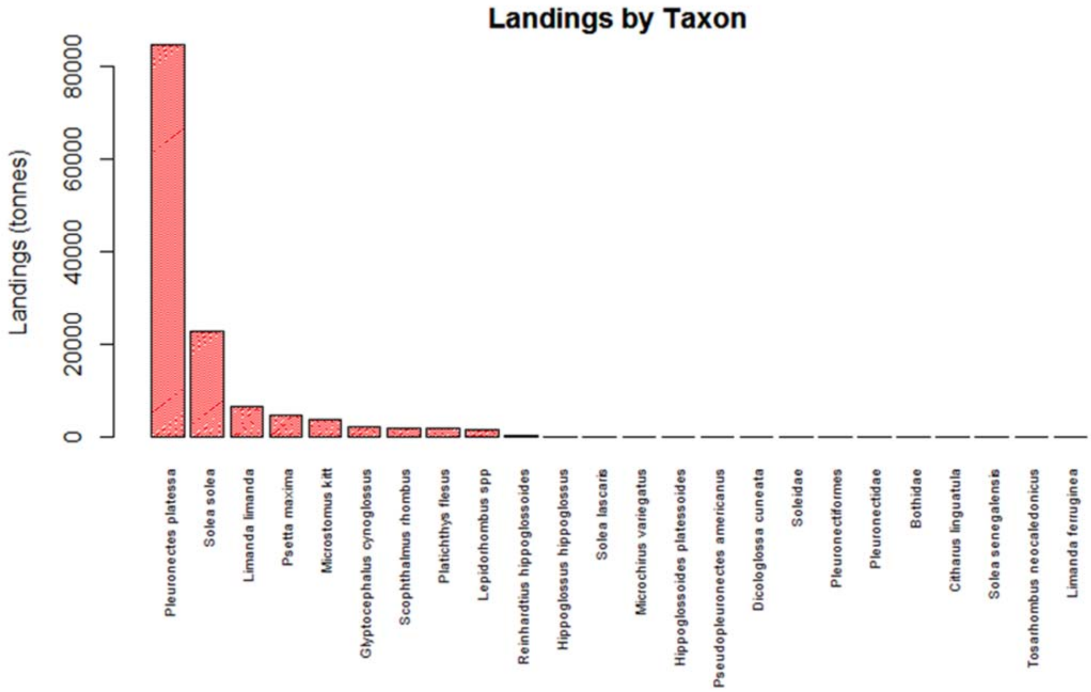


Figure 2. Total landings weights in tonnes per species (or taxon) for North Sea area (divisions III, IV and VIId) mean of 2013 and 2014. Species presented represent 95% of the total landings of flatfish species.

Landing locations

Based on total annual landings of flatfish by weight, mean of 2013 and 2014, the top 20 harbours are presented in Table 3. Most flatfish is landed in the harbour of Harlingen, situated in the north of the Netherlands. All top 10 harbours are located in the Netherlands and Denmark. Important harbours in other countries are: Oostend and Zeebrugge (Belgium), Boulogne-sur-Mer (France), Peterhead and Lerwick (Scotland).

Table 3. Landings of flatfish in weight (tonnes) mean 2013 and 2014 by harbour (top 20).

Ranking	country	harbour	Landings (weight in tonnes)
1	Netherlands	Harlingen	26988
2	Netherlands	Den Helder	13194
3	Denmark	Thyborøn	11596
4	Netherlands	IJmuiden	9478
5	Netherlands	Vlissingen	5921
6	Denmark	Hvide Sande	5229
7	Denmark	Hanstholm	5141
8	Netherlands	Eemshaven	5074
9	Netherlands	Stellendam	5013
10	Denmark	Hirsthals	4488
11	France	Boulogne-sur-Mer	3637
12	Netherlands	Lauwersoog	3604
13	Netherlands	Scheveningen	3057
14	Denmark	Thorsminde	2376
15	Belgium	Oostende	2350
16	Belgium	Zeebrugge	2223
17	Scotland	Peterhead	1549
18	France	Dunkerque	1412
19	Scotland	Lerwick	981
20	Denmark	Thorup Strand	873

Landings by fleet: vessel lengths and flag (country)

The fleet of Dutch vessels with lengths over 40 meters is clearly the most dominant in landing flatfish species (Figure 3 and Table 4).

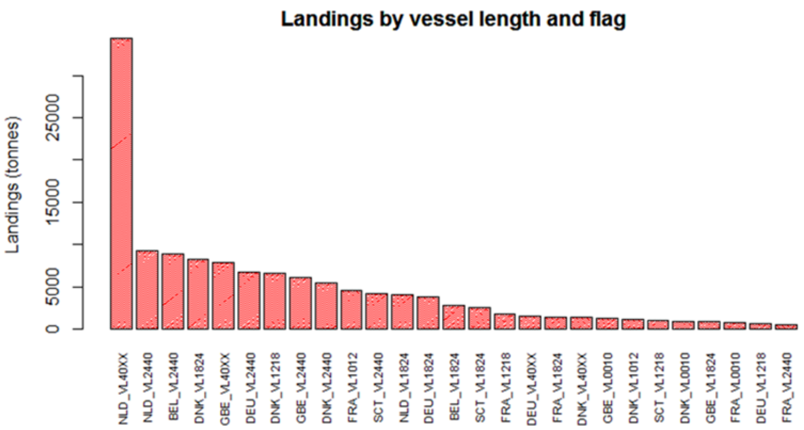


Figure 3. Annual landings of flatfish in weight (tonnes) mean of 2013 and 2014 per vessel length and flag.

Table 4. Annual landings of flatfish in weight (tonnes) mean of 2013 and 2014 per vessel length (in meters) and flag (top 10).

Flag	Vessel lengths (m)	Landings
Netherlands	≥40	34420
Netherlands	24-40	9205
Belgium	24-40	8946
Denmark	18-24	8263
England	≥40	7922
Germany	24-40	6772
Denmark	12-18	6598
England	24-40	6037
Denmark	24-40	5436
France	10-12	4575

Effort (number of trips) by fleet: vessel lengths and flag (country)

English and French fleet of vessels between under 12 meters of length conduct the most numbers of trips on an annual basis (Figure 4).

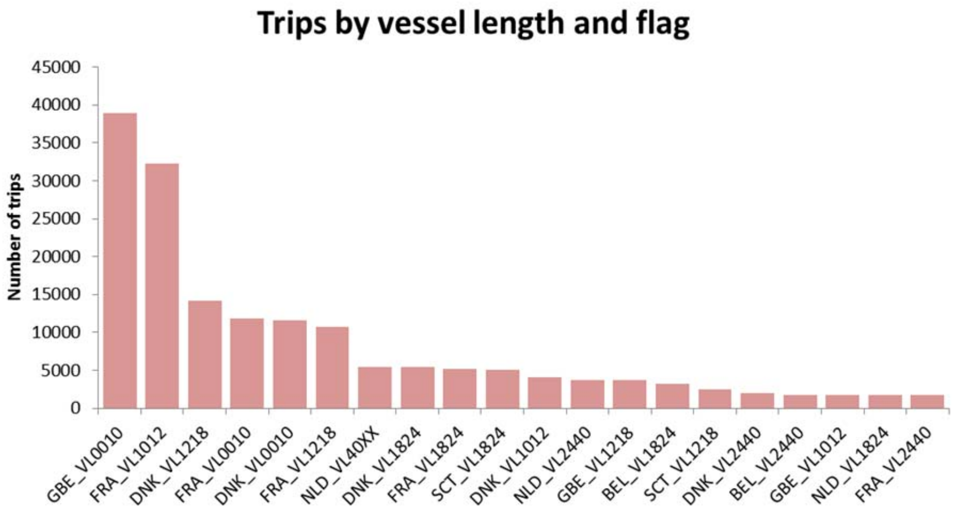


Figure 4. Number of trips for the most important flat fish fisheries (top 20) by vessel length and flag (country) sum of trips in 2013 and 2014.

Spatial distribution of landings and effort

Distribution of flatfish landings and effort per ICES rectangle for the North Sea are (ICES divisions III, IV and VIId) are presented in Figures 5 and 6.

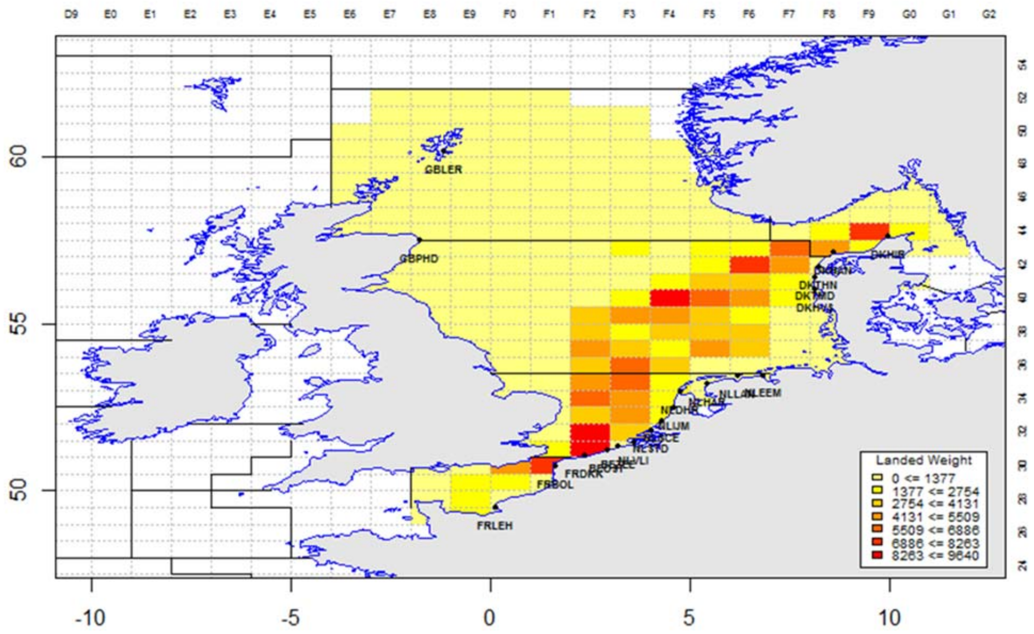


Figure 5. Distribution of annual landings (sum of 2013 and 2014) of flat fish in the North Sea area (ICES divisions III, IV and VIId).

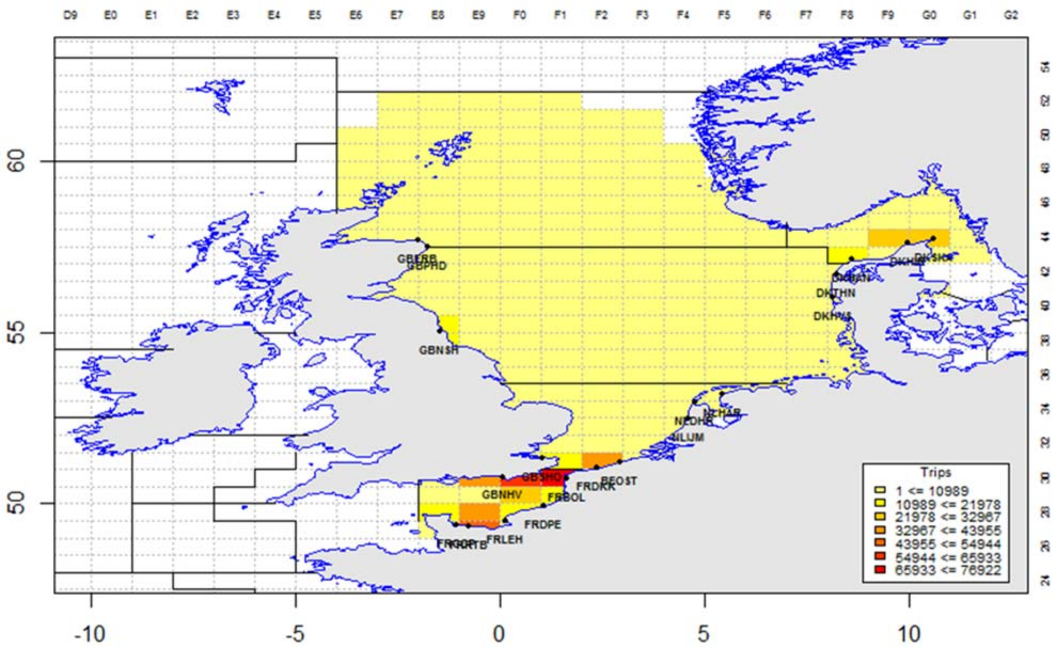


Figure 6. Effort distribution of flat fish fisheries, sum of trips of 2013 and 2014 in the North Sea area (ICES divisions III, IV and VIId).

2.6 Discard estimates

Based on the “Discard Atlas of the North Sea Fisheries” a summary is given of the landings and discards for the areas and the fisheries in the North Sea (IIIa,IV, VIId). An elaborative description can be found in http://edepot.wur.nl/315708:

Area IIIa
For the period 2010-2012 in Skagerak, discard ratios for the commercially important species plaice, cod and Norway lobster were 10%, 34 % and 41 % respectively. The relatively higher discard ratios for cod in the Skagerrak is assumed to be a result of using 90 mm trawls, which have an insufficient size selectivity in relation to minimum landing size of cod. Besides, the Skagerrak is an area with high relative abundance of juvenile cod. The main reason for Norway lobster discards is a mismatch between trawl selectivity and minimum landing size, which is 40 mm carapace length in area IIIa. Discards of Northern prawn (9%) are generally attributed to small individuals with low commercial value. Dab, a species that is not subject to catch limits in the Skagerrak, exhibit high discard ratios due to low commercial value. Large mesh otter trawls are predominantly used to catch plaice in the Skagerrak. Discard ratios are expected to be low, but are still significant for cod (27%).

Area IV
During the period 2010 to 2012 dab had by far the highest discard ratios in the North Sea, 91% on average. The high abundance of dab and the low market value contributed to this result. The second highest discard ratio was for plaice, 43% on average. In contrast to the two mentioned flatfish species, plaice and dab, discard ratios for sole were much lower (13% on average) demonstrating the high market value and the ability of fishermen to avoid unwanted by-catch of sole. Discard ratios for the more abundant lemon sole were more stable, on average 22%. Discard ratios for the high value species turbot and brill were below 5% in all years. For some by-catch species (anglerfish, megrims, Greenland halibut, blue ling, tusk) extremely low or even zero discards were reported. The flatfish fisheries with beam trawls with 80 mm mesh produced high discard ratios especially for plaice, dab and whiting. Lowest discard ratios were reported for fisheries with gillnets and large meshed beam trawls (mesh of 120 mm and above).

Area VIId
In the Eastern Channel between 10-15% of dab, plaice and lemon sole catches are being discarded. In 2010, the highest discard/catch ratio was observed for dab (64%). For many of the demersal species, including flat fish, discard/catch ratios varied significantly between years.

Description of current national sampling programmes
The annual numbers, based on 2013, of sampled market events and trips per country are presented in Table 5. Most of the countries carried out onshore sampling based on a sampling frame of market visit-trip, while Sweden, German, and Belgium conducted onshore sampling based on vessel-trip. These numbers represent current existing markets sampling programmes conducted by the different institutes in Europe covering a wide range of different commercial species, not just flat fish species.

Table 5. The annual number of sampled market events and trips per country in 2013.

Country	# market events	# sampled trips
SCT	190	494
GB	339	666
NLD	78	237
SWE (vessel-> trip)	- (57)	57
DNK	84	237
FRA	44	46
DEU (vessel-> trip)	- (7)	7
BEL (vessel-> trip)	- (33)	33
total	832	1777

2.7 Conclusions from fisheries descriptions

Plaice, *Pleuronectes platessa*, is the most abundant species in the landings for the North Sea area. Other commercially important species are sole (*Solea solea*), turbot (*Psetta maxima*) and brill (*Scophthalmus rhombus*). The top 5 of annual flat fish landings includes Dab (*Limanda limanda*), but this species is economically less valuable.

The Netherlands is the main fishing country in terms of landings and TAC allocations for flat fish. In terms of fishing effort, in number of trips, England and France are the two most important countries, due with large fleets of relatively small vessels (< 12m).

The most important flat fish harbours are situated in the Netherlands and Denmark.

3 Regional Sampling Scenarios

3.1 Data call

Scope of the data call of case study 3 Southern North Sea flatfish:

- Vessels: Vessels using demersal fishing techniques.
- Target species: Common sole (*Solea solea*), Plaice (*Pleuronectes platessa*), Turbot (*Psetta maxima*), and Brill (*Scophthalmus rhombus*).
- Countries: The Netherlands, United Kingdom (England, Scotland, Wales and Northern Ireland), Belgium, Germany, Denmark, France and Sweden.

Data was be provided in the format presented in Table 6.

Table 6. Format data call.

Name	**Comments**
RecType	on-shore scheme, at-sea scheme or both.
vslFlgCtry	Vessel flag country.
voyageId	Unique trip code.
vslId	Unique vessel id.
vslLenCls	Vessel length classes.
vslPower	<221kW,>221kW
vslType	Predominant fishing gear used.
depDate	Departure date for the trip.
arvDate	Arrival date for the trip.
homePort	Port in which the vessel is registered.
depLoc	Departure port.
arvLoc	Arrival port
sppCode	WoRMS code
Spp Name	WoRMS Scientific name
landWt	Landed weight in kg.
atSeaSampLoc	The location to access the trip.
onShoreSampLoc	Sample location.
Rect	ICES rectangle

Area	ICES subarea, division or subdivision
foCatEu6	EU level 6 métier.
fleetNat	National fleet definition.
onShoreStratNat	Stratification for on-shore sampling.
atSeaStratNat	Stratification for off-shore sampling.
domainNat1	Any current National estimation domain.
domainNat2	Any current National estimation domain.

3.2 Defining the sampling population

ICES area IIIa, IV and VIId constitutes more than 99.9% of the total landings of the flat fish species as defined in the data call. Therefore, the sampling trip population was limited to areas IIIa, IV and VIId. Step 1: Beam trawlers (TBB), demersal trawlers e.g. otter trawls and demersal seiners (DTS) and gillnetters (DFN) are the predominant vessel types to target of all 4 species (Figure 7) (accounts for 99% of the total landing). Since, these 3 vessel types are very likely to targeting flat fishes and the other types mainly record zero's for flatfish landings, it was decided to sample only these three types of vessels to test different sampling designs for flat fish.

Step 2: Trips using the following gears: dredge (gear code: DRB), fishing traps (gear codes: FOO, FPO, FYK), hooks and lines (gear codes: LHM, LHP, LLS, LTL), mid-water trawls (gear codes: OTM, PTM) , purse seines (gear code: PS) , pair bottom trawls (gear code: PTB), and other fishing gears (code: MIS) are most likely not targeting flat fishes. Excluding these trips kept 98.3%, 96.7%, 98.2% and 97.6% of the total landings of the target species after step1. Therefore, these trips with such gears were excluded in the sampling population.

Step 3: Beam trawer TBB with small mesh size (< 79 mm) are mostly targeting brown shrimp (*Crangon crangon*). Excluding these gears kept 98.2%, 96.6%, 98.0% and 97.5% of the total landing of the four flat fishes after step 1.Therefore, beam trawls with mesh size below 79 mmm were not included in the sampling population.

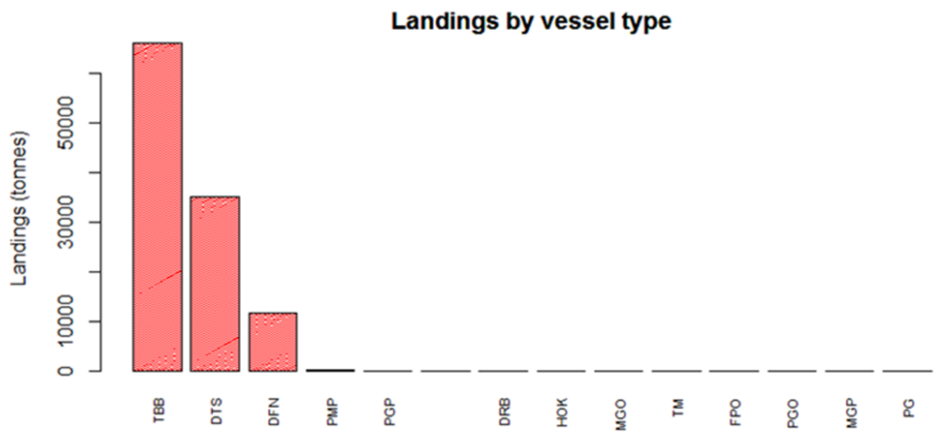


Figure 7. Landings by vessel type. Beam trawl (TBB), demersal trawlers (DTS) and gillnetters (DFN) are the important contributors to flatfish landings in the North Sea.

3.3 Simulations of sampling scenarios

Simulations were conducted on 2013 data, which, after defining of the sampling population, included information of 121199 trips. The annual numbers of sampled market events and trips per country are presented in Table 5. In total 1777 trips were sampled. For ease of simulation, we treat these numbers all as market visit – trip and use 1800 trips in the sampling simulation.

Simulations were conducted on several scenarios which were relevant for flat fish fisheries. The goal is to provide several examples on how to choose a good sampling design. For simplification, all scenarios were conducted for plaice landings. For each simulation, a random sample is selected (with 1800 trips in the end) according to a specific sampling design, and the total weight is estimated per domain (i.e. area and quarter) using the Horvitz-Thompson estimator. The process was repeated for 2000 times (n= 2000). Eventually, the performance of the sampling design can be judged by the mean and the variance of the estimate from the n replicates. For each scenario, the sampling designs were judged according to two aspects, bias and precision: 1) A good sampling design should be unbiased. Consequently, the mean of the estimated total should be the same as the true population value. 2) A good sampling design should achieve low variance (high precision) of the estimated total. A ratio of variance of <1 between a sampling design and a reference sampling design (i.e. design effect) indicates an improved sampling design. Additionally, due to the inappropriate strata and sampling effort allocation, some features in the sample might not be representative of the population; sometimes the expected strata might not be accessible during the sampling stage. Therefore, a post-stratification step is applied in the simulation to obtain the unbiased estimate for each domain.

The simulation procedure is as follows:

1. For run number *k* (*k*=1,…2000) Repeat steps 1.1 to 1.X 2000 times
   1.1. Select a sample from the trip population according to a specific sampling design *i, i* refers to one of the sampling design in the following scenarios;

1.2. Calculate the sample inclusion probability or weight according to the sampling design *i* (i.e. the strata or cluster);

1.3. Adjust the inclusion probability or weight according to the output domain (i.e. area and quarter);

1.4. Estimate the total weight per domain using the adjusted probability or weight. Steps 1.2 to 1.4 are the Horvitz-Thompson estimator with post-stratification;

1.5. Record the estimated total weight for domain j (j refers to the domains) and run *k*: $totweight_{ijk}$

2. Repeat steps 1.1 to 1.5 for 2000 times, and compute the estimated total weight for design *i* domain *j* as: $\overline{totweight_{ij}} = (\sum_{k=1}^{2000} totweight_{ijk})/2000$;
3. The bias of design *i* can be judged by comparing $\overline{totweight_{ij}}$ with the total weight computed from the population data ($TOTWEIGHT_{ij}$);
4. Compute the estimated variance of total weight for design *i* domain *j* as: $\widehat{\mathrm{var}}(\overline{totweight_{ij}}) = \mathrm{var}(totweight_{ijk})$;
5. Compute the design effect of the current sampling design i compared to the reference sampling design as : $\widehat{\mathrm{var}}(\overline{totweight_{i,j}}) / \widehat{\mathrm{var}}(\overline{totweight_{i=\mathrm{ref},j}})$

3.3.1 Scenario 1

In a common regional sampling programme, the sampling estimates are often presented per area and quarter (so called domain). As benchmark we applied *a one-stage sampling with trip as the sampling unit* (i.e. a simple random sampling design) with post stratification per domain. Stratification can be applied to increase precision in a sampling design. In this scenario, we investigated the possibility of applying a stratified sampling, using vessel length as strata (<18m, 18m-40m, ≥40m). Two sampling effort allocation methods were applied – proportional and Neyman. Therefore, we compare:

a. Reference: simple random sampling of trips; estimated total weight post-stratified by quarter + area (SRS-poststrat_area+quarter).
b. Stratified with proportional allocation: number of sampled trips per stratum is proportional to the number of trips per stratum; estimated total weight post-stratified by quarter + area (Strat_Prop_vsllen_poststrat_area+quarter).
c. Stratified with Neyman allocation: number of sampled trips per stratum is proportional to the number of trips per stratum and its variance; estimated total weight post-stratified by quarter + area (Strat_Neym_vsllen_poststrat_area+quarter).

The original strata by definition should be vessel engine power, however due to the lack of a complete data set, vessel length was used as a surrogate.

Stratification with effort proportional to the number of trips reduced variance only slightly, by around 0-10% in most domains (Table 7), whilst using the Neyman allocation reduced variance substantially in all but one domain, by around 30-60%, with the biggest improvement in area IV, from which the majority of the landings are taken (Table 7). All three scenarios produced unbiased estimates (Figures 8, 9 and 10).

Table 7. Ratio of variance between a sampling design and a reference sampling design indicating the performance of the sampling scenarios. A ratio <1 indicates an improvement in design.

Domain (area-quarter)	**Scenario-1b/ reference (scenario-1a)**	**Scenario-1c/ reference (scenario-1a)**
IIIa-Q1	0.89	0.66
IIIa-Q2	0.96	0.72
IIIa-Q3	0.98	0.64
IIIa-Q4	1.07	0.65
IV-Q1	0.95	0.59
IV-Q2	0.95	0.41
IV-Q3	0.95	0.44
IV-Q4	0.92	0.44
VII-Q1	1.03	0.51
VII-Q2	0.99	1.80
VII-Q3	0.97	0.63
VII-Q4	1.01	0.48
Total	**0.95**	**0.46**

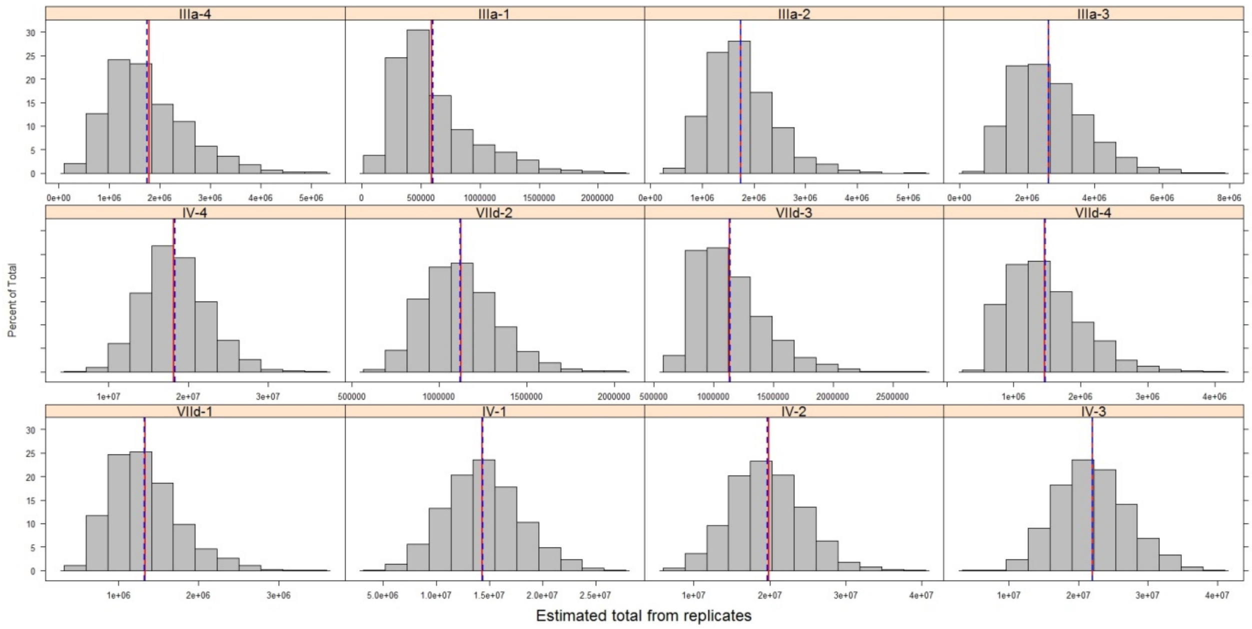


Figure 8. Scenario 1a. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

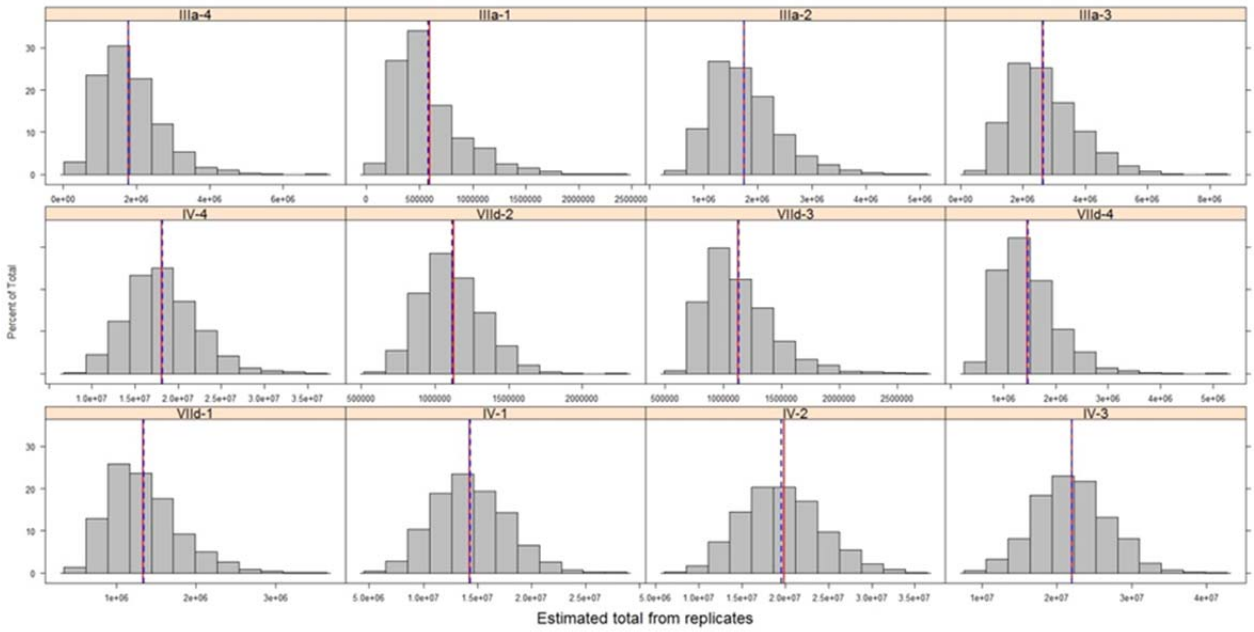


Figure 9. Scenario 1b. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

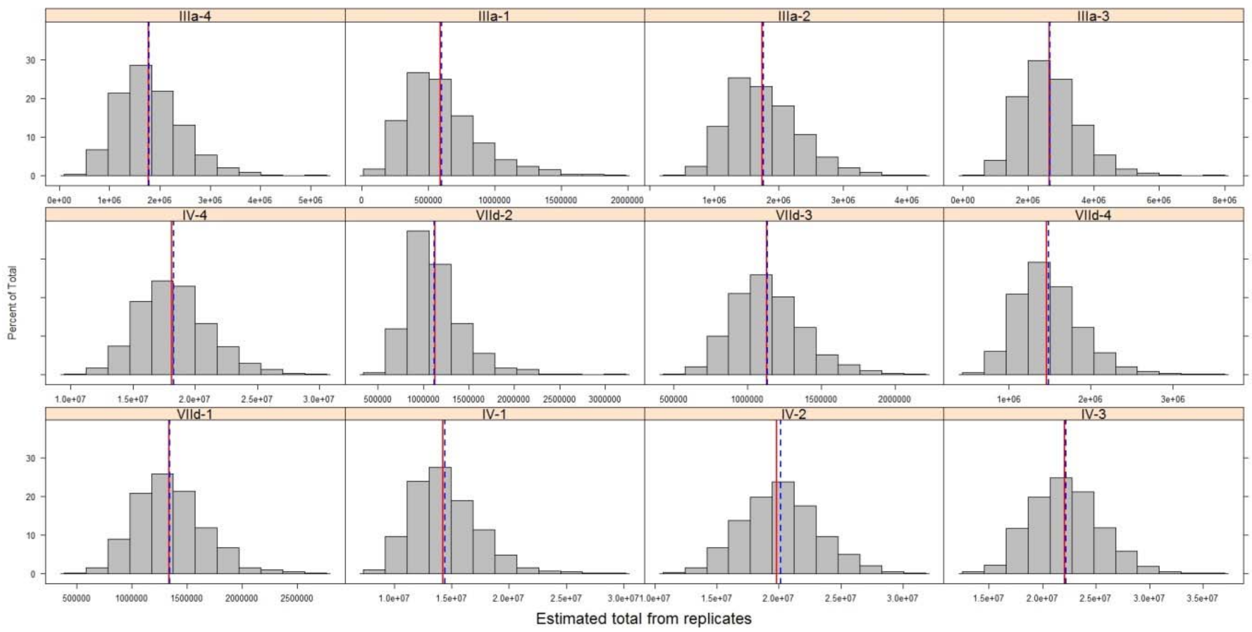


Figure 10. Scenario 1c. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

3.3.2 Scenario 2

To investigate the possibility of applying a stratified sampling, using 'vessel length-area' and 'vessel length-quarter' as strata. Neyman allocation, post-stratified by area or quarter. Therefore, we compare:

a. Reference: simple random sampling of trips; estimated total weight post-stratified by quarter + area (= Scenario 1a).
b. Stratified by quarter + vessel length. Neyman allocation, post-stratified by area.
c. Stratified by area + vessel length. Neyman allocation, post-stratified by quarter.

Stratification by quarter, using Neyman effort allocation, with post-stratification by area, reduced the variance in almost all domains apart from VIId-Q2 by around 60% (Table 8). Stratification by area, using Neyman effort allocation, with post-stratification by quarter, substantially reduced variance in area IV by around 75%, but this was at the cost of substantially increased variance in VIId, and little effect on variance in IIIa (Table 8). This is because most of the sampling effort has been allocated to IV. The estimates were unbiased (Figures 11 and 12).

Table 8. Ratio of variance between a sampling design and a reference sampling design indicating the performance of the sampling scenarios. A ratio <1 indicates an improvement in design.

Domain (area-quarter)	**Scenario-2b/ reference (scenario-2a)**	**Scenario-2c/ reference (scenario-2a)**
IIIa-Q1	0.52	0.90
IIIa-Q2	0.83	1.03
IIIa-Q3	0.67	0.90
IIIa-Q4	0.75	1.09
IV-Q1	0.43	0.33
IV-Q2	0.38	0.23
IV-Q3	0.39	0.25
IV-Q4	0.40	0.24
VII-Q1	0.49	2.10
VII-Q2	2.44	6.89
VII-Q3	0.62	2.37
VII-Q4	0.48	1.88
Total	**0.40**	**0.29**

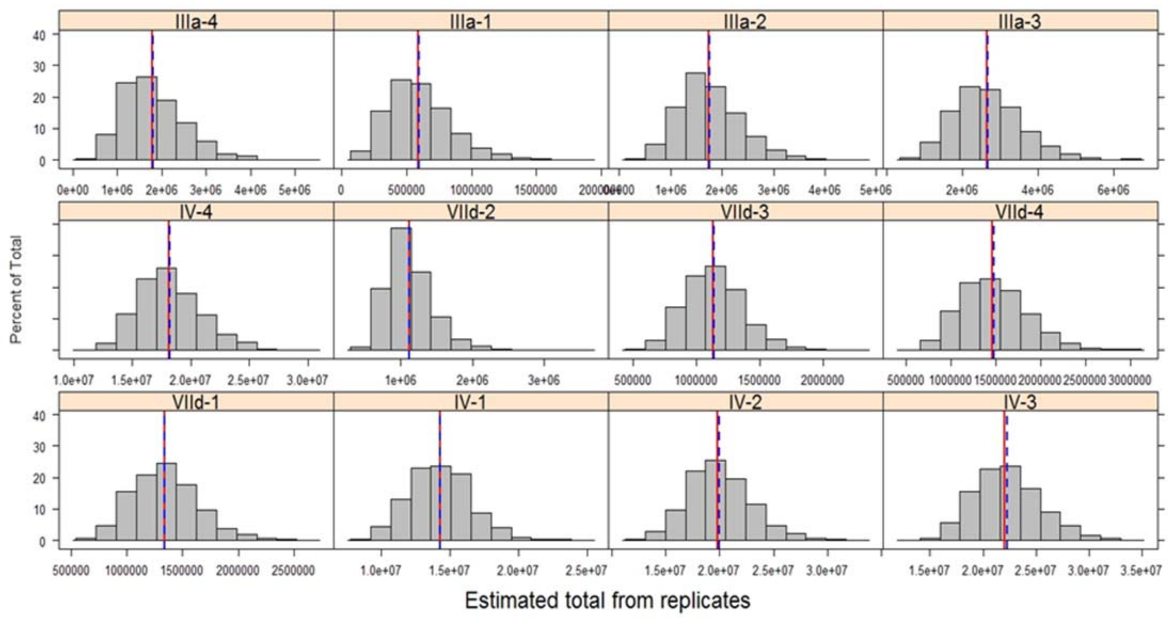


Figure 11. Scenario 2b. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

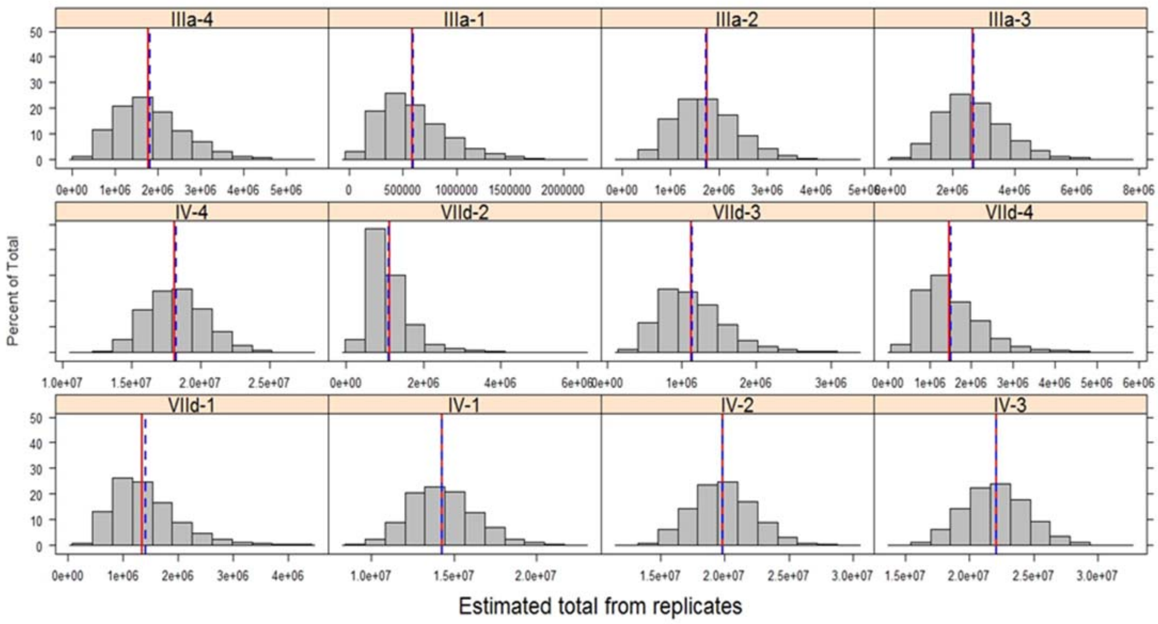


Figure 12. Scenario 2c. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

3.3.3 Scenario 3

A stratified two-stage clustered sampling (primary sampling unit = landing harbour, secondary sampling unit = trip) design was tested: strata = quarter + vessel_length, cluster = landing harbour sampling with replacement, sample 1 trip per harbour without replacement. Number of harbours sampled per strata is proportional allocated by the number of trips, post-stratification to area_quarter. These different scenarios were compared:

a. Reference: simple random sampling of trips; estimated total weight post-stratified by quarter + area (= Scenario 1a).

b. Sampling harbours with equal probability, simple random sampling, within each strata.

c. Sampling harbours with probability proportional to number of trips, within each strata.

The two-stage cluster sampling, where one trip was sampled from each harbour increased the variance substantially, by more than 3 times overall, with similar results for both scenarios (Table 9). The estimates were unbiased (Figures 13 and 14).

Table 9. Ratio of variance between a sampling design and a reference sampling design indicating the performance of the sampling scenarios. A ratio <1 indicates an improvement in design.

Domain (area-quarter)	Scenario-3b/ reference (scenario-3a)	Scenario-3c/ reference (scenario-3a)
IIIa-Q1	2.48	1.95
IIIa-Q2	5.36	6.03
IIIa-Q3	3.34	4.49
IIIa-Q4	4.09	3.64
IV-Q1	2.42	2.64
IV-Q2	3.81	3.96
IV-Q3	3.49	3.94
IV-Q4	3.17	3.83

VII-Q1	2.98	2.60
VII-Q2	9.59	6.18
VII-Q3	3.21	3.57
VII-Q4	2.99	4.49
Total	**3.35**	**3.73**

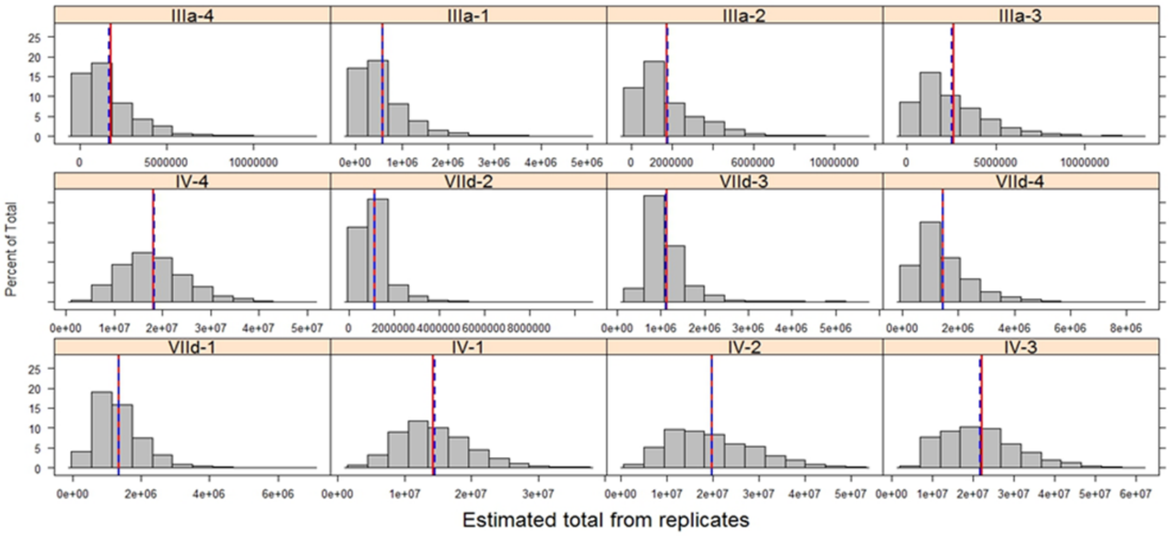


Figure 13. Scenario 3b. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

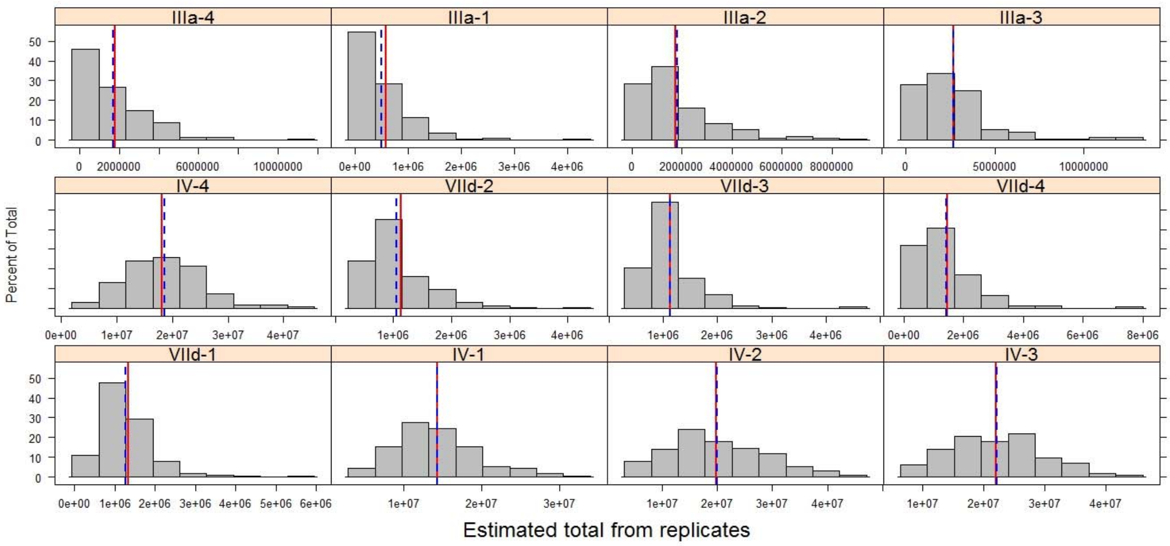


Figure 14. Scenario 3c. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

3.3.4 Scenario 4

In this scenario we used stratified two-stage clustered sampling (strata = quarter + vessel_length, cluster= landing harbour), while the number of harbours sampled per strata is according to proportional allocation of number of trips and the final total weight is post-stratified into area+quarter. Currently sampling is done through national on-shore and at-sea monitoring programmes, countries only sample vessels of their own nationality. Figure 15 presents the distribution of landings of plaice per country in different harbours. To see the effect of a national programme compared to a regional or international approach in sampling, we compare:

a. Reference: Sampling from all trips landing in the harbour
b. Sampling from trips belonging to the same country flag of the harbour (proxy for national sampling scheme).

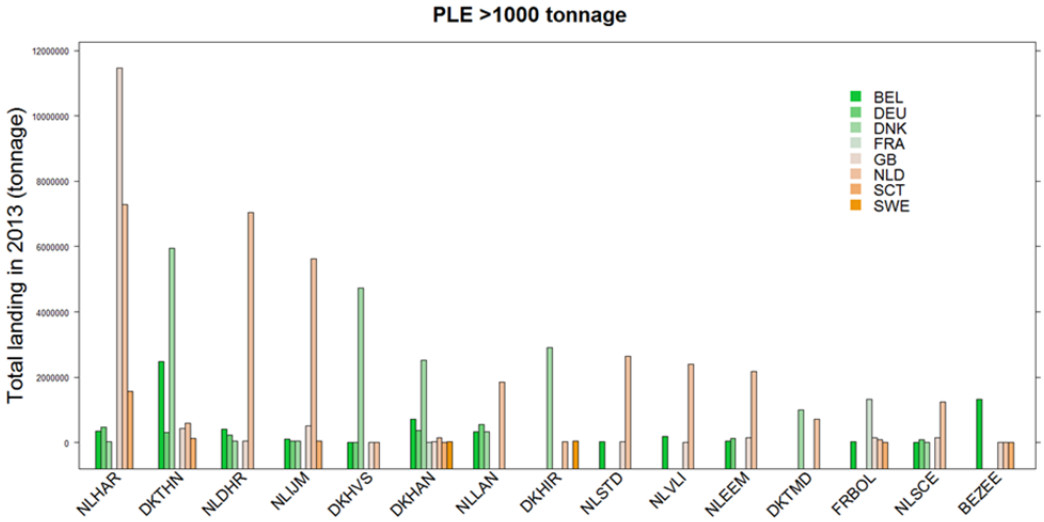


Figure 15. Landing composition of vessel flags per landing location (harbour) for PLE with >1000 tonnage, in decreasing order.

The regional sampling scheme gave unbiased estimates (Figure 15) but the sampling scheme based on country flag, the proxy for national sampling scheme, gives biased estimates of the total weight (Figure 16).

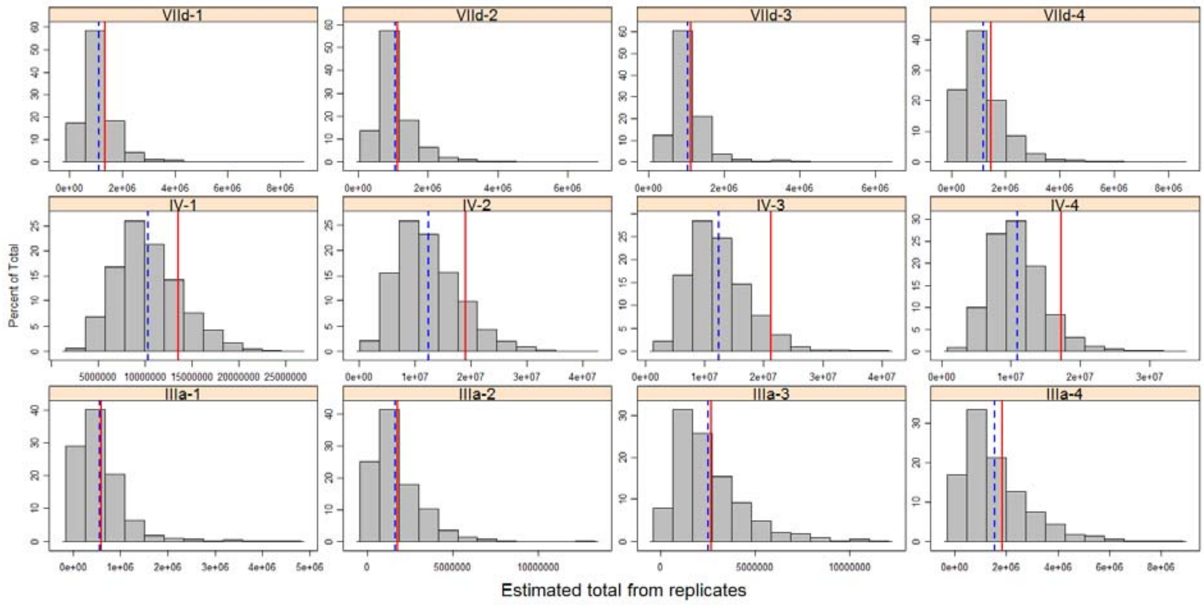


Figure 16. Scenario 4b. Distribution of estimated total from replicates (n= 2000). Dotted line is the real mean, red line is mean from simulations.

3.4 Conclusions

This simulation study is the first attempt to conduct a statistically sound design based regional sampling programme on flat fish. Usually the fisheries sampling are conducted either on-shore or at sea. In this simulation, we focused on the on-shore sampling, where market-visits and trips are the main sampling frames.

A population data set of all fishing trips from the participating countries were collected. Most important is the knowledge we gained in re-defining the sampling population. By 1) limiting the vessel types to beam trawlers (TBB), demersal trawlers e.g. otter trawls and demersal seiners (DTS) and gillnetters (DFN), 2) limiting the gear usage of the trip and 3) excluding TBB with small mesh size (< 79 mm), we managed to retain more than 95% of the total flat fish landings while excluding a large number of trips with zero catches. This significantly decreases the sampling variance. The excluded vessels or trips could apply a different sampling strategy, such as self-sampling.

Stratification is used in sampling to increase precision while keeping an unbiased estimate. Therefore, we applied stratification in scenario 1 and 2, with several strata (vessel-length, quarter+ vessel-length, area+ vessel-length) and 2 sampling effort allocation principles (proportional and Neyman). Results show that quarter+ vessel-length with Neyman allocation gives the best sampling performance (lowest sampling variance) compared to a simple random sampling scheme.

Clustering is used in sampling to reduce the logistic cost, or increase feasibility, but usually at a sacrifice of lower precision. Therefore, we applied clustered sampling + stratification in scenario 3, with 2 sampling effort allocation principles (equal probability, probability proportional to size). Results show that the stratified two-stage sampling reduces the precision.

In a regional sampling scheme, the landing composition of a harbour usually contains vessels from multiple countries. Therefore, the biggest advantage of a regional sampling is to sample all the international trips, rather than the trips belonging to the country of the landing harbour (which is usually conducted at national sampling scheme). Therefore, we compared these two sampling schemes in scenario 4. Results show that a regional sampling scheme gives unbiased estimate, while the estimate from the national scheme is biased.

The selected optimal sampling scheme is in the case study was stratified by quarter and vessel length with Neyman allocation, post-stratified by area (scenario 2b). Based on this scenario and the optimal sampling effort distribution per country is calculated. Table 10 presents the current distribution of sampling effort between countries, based on the different national sampling programmes (1777 trips sampled) and the distribution of sampling effort between countries based on scenario 2b (1800 trips sampled). A transfer to a regional sampling scheme based scenario 2b would imply a significant change in numbers of trips sampled per country, however the current national sampling programmes are not based on flat fish sampling only, as is the situation in the case study.

Table 10. Current distribution of sampling effort and distribution based on the optimal sampling scheme in the case study (scenario 2b). Sampling effort is the annual number of trips sampled.

Country	Current sampling effort	Based on scenario 2b
BEL	33	84
DEU	7	46
DNK	237	367
FRA	46	316
GB	666	238
NLD	237	365
SCT	494	231
SWE	57	153

ANNEX 14 WP2 CS4 Southern and Northern hake stocks sampling designs

The fishPi project
Regional co-operation in fisheries data collection

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP2.3 – Case Study fisheries

Deliverable CS4 – Hake Southern and Northern stocks sampling design simulations

Lead: IEO, Spain

Core Team: José Castro, José Rodríguez, Eva Velasco (IEO)
João Pereira, Nuno Prista, Marina Dias, Manuela Azevedo (IPMA)
Lucia Zarauz, Estanis Mugerza, Marina Santurtún (AZTI)
Michel Betignac, Laurent Dubroca (IFREMER)
Hans Gerritsen (MI)
Liz Clarke (MSS), Alastair Pout (MSS)

Data providers: IEO (Spain), AZTI (Basque Country, Spain), IMARES (Netherlands), IPMA (Portugal), IFREMER (France), MI (Ireland), CEFAS (United Kingdom), ILVO (Belgium), MSS (United Kingdom), DTU-Aqua (Denmark), SLU (Sweden).

Table of Contents

1 Data

The data needed to develop the CS4 simulations follow the guidelines established by the WP2 team for all case studies.

1.1 Data request

The Data Sharing Agreement was made effective on 10/04/2015.
A specific fishPi data format, composed of 25 fields, was decided by the fishPi team (Table 1.a) and sent to the National data suppliers under a formal data request data on 02/06/2015.
Scope of the CS4 data call:

- Vessels: Vessels using demersal fishing techniques (mainly trawls, set gillnets and set longlines).
- Areas: Division IIIa, Subareas IV, VI, VII and VIII, and Division IXa.
- Target species: hake (*Merluccius merluccius*).
- Years: 2013 and 2014.
- Data submitters: countries with hake TAC.

Despite having initially requested the fishing activity of the demersal fleet, during the WP2 coordination meeting (Aberdeen, 9th-10th June 2015) it was decided to extend the requested data to the total fishing activity developed by each country in the study area, including trips with and without hake landings. Most of data supplier countries sent their respective fishPi files during the summer 2015, but the final matrix could not be completed until the full provision of the latter country in November 2015.

1.2 fishPi CS4 data matrices

Once the respective national data sets were compiled in one CS4 matrix, some decisions had to be taken. Due to problems with the French data submission, which displayed only positive trips, *i.e.* trips with landings of hake, two fishPi CS4 data sets had to be considered:

- CS4-total data set: all countries exploiting hake. However, as France only provided positive trips of hake, the same criteria were to be applied to the trips of the remaining countries in order to harmonize data.
- CS4-shke data set: data set containing only trips developed in Iberian waters (ICES Divisions VIIIc and IXa), where the Southern stock of hake is distributed. This data set includes all the trips developed in the area by Spanish and Portuguese vessels, including trips with no hake, and respects the original approach agreed by the WP2 team to apply in all case studies.

In both sets of data, field 17 ("onShoreSampLoc") was covered with the United Nations Code for Trade and Transport Locations (UNLOCODE) of the landing port or the sale market which facilitates access to the sample. This field is useful to indicate the country responsible for the sampling, which is fundamental to organize a regional sampling program.
The last two fields, reserved to indicate the National estimation domains, were covered differently in each dataset:

- Field 24 ("domainNat1"): originally this field had been reserved to replicate the current post-stratification used to provide commercial data to ICES, *i.e.* a combination of technical (métier), spatial (ICES Division), and time (quarter) disaggregation. However, the huge number of métiers developed in North Sea and Western Atlantic waters made it necessary to reduce this filed to a combination of Division*quarter in the CS4-total data set. In the case of CS4-shke data set, the Iberian subareas used by the Spanish (AZTI and IEO) and

Portuguese (IPMA) scientific institutes for sampling of the coastal fishing activity were used (Table 1.b).

- Field 25 ("domainNat2"): this field was used only in the CS4-total data set to differentiate both hake stocks.

The work developed in CS4 under the fishPi project is an opportunity to use the Iberian subareas established by the Iberian institutes for their respective National sampling programs together. Once they are used in a broader context beyond the national scope, it has been observed that the coding should be improved in the case of moving towards a regional sampling design.

1.3 Sampling coverage by country

Sampling coverage by country ("magic numbers") indicates the annual number of sampled trips by country for year 2014 (Table 1.c). Two numbers were requested, the total number of trips sampled at sea and the number of trips sampled onshore. Besides the total number of sampled trips, the number with occurrence of hake was also requested.

1.4 Data setting

Data analysis to inform scenarios was based on year 2013.

1.4.1 **CS4-total data setting**

The year 2013 of the CS4-total data set was used to set some extra fields to facilitate the stratifications to be used to run the sampling design simulations: stratifications by port, country and institute were performed.

- Stratification by Country: stratification by country was developed by using Field 17 ("onShoreSampLoc") extracting the country code from the first two letters of the harbour code (UNLOCODE). This information yields the same number of countries, but with different allocation of trips and tons of hake.
- Stratification by Institute: a subdivision of country stratification was performed for United Kingdom and Spain.
  - United Kingdom (GBR) was subdivided in order to differentiate Scottish sampling (developed by MSS institute) from the remainder UK sampling (developed by CEFAS).
  - Trips landed in Spain (ESP) were subdivided to distinguish between trips landed in Basque ports, which are sampled by AZTI, and the remainder Spanish ports which are sampled by IEO. Both Spanish Institutes have access to landings of both hake stocks (Figure 1.a).
- Stratification by port: two port strata were made, one for the main ports with landings higher than 1000 tons of hake ("major") and one for the secondary ports ("minor"). The first one contains 70% of overall hake landings, but this proportion varies among countries.

1.4.2 **CS4-shke data setting**

Similarly, year 2013 of the CS4-shke data set was used to set some extra fields to facilitate the stratifications to be used to run the sampling design simulations. In addition to the quarterly stratification, which is easily derived from the date of arrival to port, stratifications by port and country were also performed.

- Stratification by Country: stratification by country was developed by using Field 17 ("onShoreSampLoc") extracting the country code from the first two letters of the harbour code (UNLOCODE). This information yields the same number of countries, but with different allocation of trips and tons of hake.

- Stratification by port: port strata were defined as the individual ports that accounted for at least 75% of hake landings. A supplementary stratum encompassed all remaining ports into a single “minor ports” stratum. Spain was subdivided to distinguish between trips landed in Basque ports, which are sampled by AZTI, and the remainder Spanish ports which are sampled by IEO.
- Stratification by port and quarter: Port*quarter strata were defined as the individual port*quarters that accounted for at least 75% of hake landings. A supplementary stratum encompassed all remaining port*quarters into a single “minor ports” stratum.

Table 1.a. fishPi data format. National data were provided in the following format as an R data frame or a comma separated CSV flat file consisting of 25 columns.

Field	Name	Comments
1	RecType	Onshore scheme, at-sea scheme or both
2	vslFlgCtry	Vessel flag country
3	voyageId	Unique trip code
4	vslId	Unique vessel ID
5	vslLenCls	Vessel length classes
6	vslPower	<221kW,>221kW (only for beam trawl fleets)
7	vslType	Predominant fishing gear used
8	depDate	Departure date for the trip
9	arvDate	Arrival date for the trip
10	homePort	Port in which the vessel is registered
11	depLoc	Departure port
12	arvLoc	Arrival port
13	sppCode	WoRMS code
14	Spp Name	WoRMS Scientific name
15	landWt	Landed weight in kg
16	atSeaSampLoc	The location to access the trip
17	onShoreSampLoc	Sample location
18	Rect	ICES rectangle
19	Area	ICES subarea, division or subdivision
20	foCatEu6	EU level 6 métier
21	fleetNat	National fleet definition
22	onShoreStratNat	Stratification for onshore sampling
23	atSeaStratNat	Stratification for offshore sampling
24	domainNat1	Any current National estimation domain
25	domainNat2	Any current National estimation domain

Table 1.b. Iberian subareas used to populate Field 24 (“domainNat1”) in CS4-shke data set.

Iberian areas	Country	Description
8cE1	ESP	ICES Division VIIIc corresponding with Asturias and Cantabrian waters
8cE2	ESP	ICES Division VIIIc corresponding with Basque Country
8cW	ESP	ICES Division VIIIc corresponding with Galician waters
9aC	ESP	ICES Division IXa corresponding with Portuguese waters for Spanish vessels
9aN	ESP	ICES Division IXa corresponding with Galician waters
9aS	ESP	ICES Division IXa corresponding with Andalusian waters
FOREIGN	PRT	ICES Division IXa corresponding to Portuguese vessels when landing in Spanish ports
NW	PRT	ICES Division IXa corresponding with Northern Portuguese ports
S	PRT	ICES Division IXa corresponding with Southern Portuguese ports
SW	PRT	ICES Division IXa corresponding with South-western Portuguese ports

Table 1.c. CS4 magic numbers. Sampling coverage by country.

Country code	Country/Institute name	Number of trips sampled at-sea	Number of trips sampled onshore	Total number of trips sampled	Total number of trips where hake was sampled
BEL	Belgium	33	--	33	17
DNK	Denmark	130	237	367	113
ESP	Spain/IEO	110	1598	1877	765
	Spain/AZTI	25	1004	1029	43
FRA	France	983	645	1628	597
GBR	United Kingdom/CEFAS	222	1711	1933	431
	United Kingdom/MSS	60	770	830	45
IRL	Ireland	67	269	336	82
NLD	Netherland	170	369	539	38
PRT	Portugal	144	1409	1553	700
SWE	Sweden	57	0	57	27
TOTAL	TOTAL	1976	7177	9153	2858

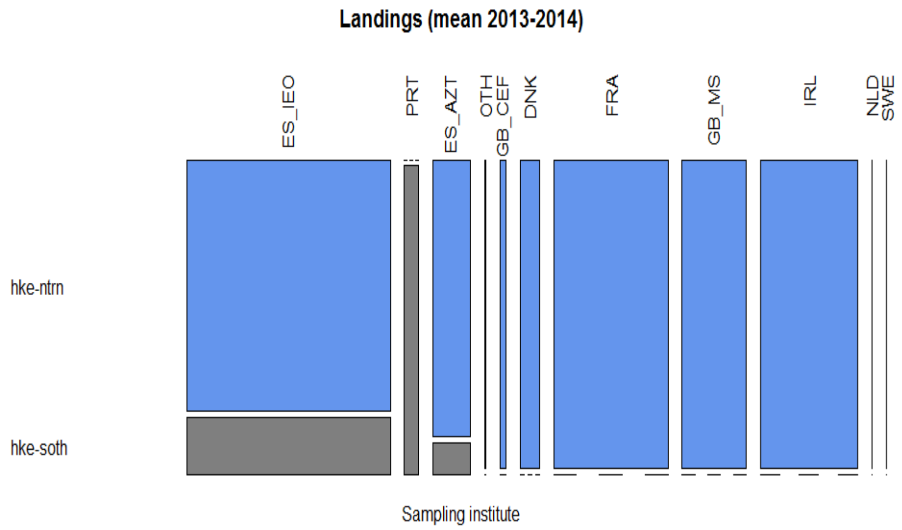


Figure 1.a. Average annual hake landings (2013-2014) by sampling country/institute.

2 Methodology

The two data sets described in Section 1 (CS4-total and CS4-shke) were used to analyse the design effect on statistical parameters in terms of bias and precision. The first step was to agree on a list of scenarios that pick up the most relevant design variations to be tested. Thus, different stratification criteria were employed under the general approach of a regional sampling program or a set of national sampling programs. Finally, the simulation runs were developed on year 2014.

2.1 Scenarios list

The list of scenarios finally decided had to be adapted to the characteristics of each CS4 data set.

2.1.1 CS4-total scenarios list

The scenarios for CS4-total data set include a 2-stage Random Sampling to be used as baseline scenario. All scenarios were performed under 2-stage clustering, with random selection of the samples without replacement, and taking a combination of market/port and day as PSU (Primary Sampling Unit), while trip was used as SSU (Secondary Sampling unit). The number of trips sampled per market event and day was set to one. Simulations were run on the 2014 data set and thus only market*days presenting landings in 2014 were considered for selection.

1. Baseline scenario: 2-stage Random Sampling without replacement (market*day as PSU and trip as SSU).
2. Stratified by country (8 strata): France (FRA), Spain (ESP), Denmark (DNK), Portugal (PRT), United Kingdom (GBR), Ireland (IRL), The Netherlands (NLD) and Sweden (SWE). The allocation of sampling effort according to the current National sampling coverage (magic numbers).
3. Stratified by institute (10 strata): two countries (GBR and ESP) were subdivided in order to reproduce the current sampling allocation between National institutes: GBR was split into Marine Laboratory (Scotland) and CEFAS (England, Wales and North Ireland), and Spain was split into Basque Country ports (covered by AZTI) and the rest of Spanish ports (IEO). Allocation of sampling effort according to magic numbers.
4. Stratified by port (2 strata): two main port strata were preformed, one with major ports and another with minor ports. Allocation of sampling effort proportional to landings.
5. Stratified by country and port: a combination of scenarios 2 and 4. Allocation of sampling effort according to magic numbers (samples by country) and proportional to landings for port strata within each country.
6. Stratified by Institute and port: a combination of scenarios 3 and 4. Allocation of sampling effort according to magic numbers (samples by institute) and proportional to landings for port strata within each institute.

In scenario 4, which is the only regional scenario considered, samples are allocated between port strata proportional to landings. The remainder scenarios reproduce the current National sampling coverage. Therefore scenario 4 with port stratification alters the current coverage between countries, so that countries with lower landings (e.g. Portugal or Sweden) have a lower sampling effort.

2.1.2 CS4-shke scenarios list

The scenarios for CS4-shke matrix include a SRS (2-stage Random Sampling) scenario to be used as baseline scenario. All scenarios were performed with replacement and 2-stage clustering, taking a combination of market/port and day as PSU (Primary Sampling Unit), while trip was used as SSU (Secondary Sampling unit). Because a significant amount of landings is registered in weekends, all days of the year were included for selection, regardless of whether or not they registered landings in 2014. In all scenarios but scenario 3d sample allocation to the different port/quarter/port*quarter strata was allocated proportional to landings (see example in Figure 2.a).

1) Base line: scenario 2-stage cluster Random Sampling with replacement and (port*day as PSU and trips as SSU).
2) Scenarios REGIONAL.
   a. REGIONAL stratified by port.
   b. REGIONAL stratified by quarter.
   c. REGIONAL stratified by port-quarter.
3) Scenario COUNTRY (country stratification with samples similar to present).
   a. COUNTRY stratified by port.
   b. COUNTRY stratified by quarter.
   c. COUNTRY stratified by port-quarter.
   d. Ctr30 (stratification with samples allocated proportional to landings).

2.2 Methodology

Regardless of the information collected in Field 1 ("RecType"), all were assumed to be sampled onshore. This assumption was made to avoid complicating the sampling designs and facilitate the interpretation of the results of the different scenarios. Obviously, this simplification should be taken into account when considering a feasibility study.
As indicated above, scenarios 1 (2-stage random sampling of trips without stratification) were used as a base line for comparison among results within each subset of data (CS4-total and CS4-shke). For each scenario, sampling designs are judged according to the following two aspects:

- Bias: ratio between the estimated values (tons of hake) and the true value registered in 2014.
- Precision: ratio between the standard error (SE) of the estimated values by each scenario in relation to the SE of scenario 1 estimates. Thus, ratios <1 indicates an improved sampling design. For CS4-shke the inter-quartile range was also computed.

In addition to these global parameters, other parameters were also calculated by stratum or domain to explore the quality of each sampling design in more detail. Particularly for CS4-shke, other parameters were calculated as estimated landings, number of PSUs, number of SSU, number of duplicated SSU (due to sampling with replacement), number of SSU with/without hake, and tonnage of hake available for sampling in sampled PSUs and SSUs by country, Iberian subarea, port, quarter, domainNat1, fleetNat, and foCatEu6.

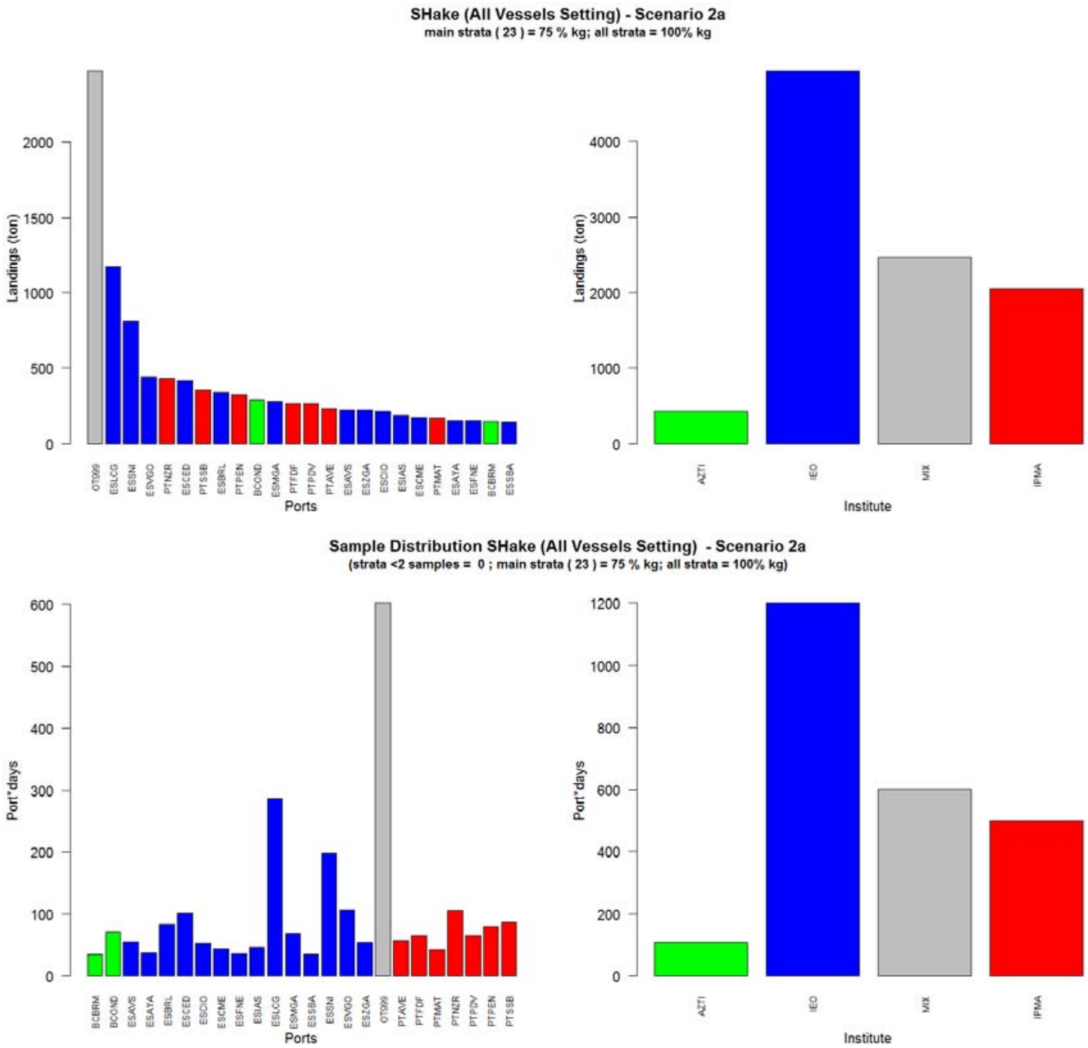


Figure 2.a. Distribution of landings across the different strata of Scenario CS4-Shke 2a (regional, stratified by port). Upper left: landings in each regional port strata. Upper right, landings by sampling institute (MIX = sampling to be carried out by several institutes depending on random draw). Lower left: distribution of PSUs. Lower right: distribution of PSUs by sampling institute (MIX = sampling to be carried out by several institutes depending on random draw).

3 Results

Following the methodology described above, the estimated landings and the respective SE was calculated by scenario for both CS4 data sets.

3.1.1 CS4-total simulation results

In Table 3.a, comparison between sampling design scenarios for CS4-total provides very similar values for bias (between 0.99 and 1.00), but clearly highest precision for scenario 4 (regional sampling design stratified by major and minor ports) and scenario 5 (stratified by country and port), as illustrated in Figure 3.a.

The analysis of coverage by “domainNat1” (combination of ICES Division and quarter) has shown no significant deviations (the respective plot is not included because of their complexity). The same analysis by “domainNat2” shows how sampling design performs similarly by hake stock (Figure 3.b).

3.1.2 CS4-shke simulation results

In table 3.b, comparison between sampling design scenarios for CS4-shke provides very similar values for bias as expected (between 0.99 and 1.01), but highest precision for scenario 2a (Regional stratified by port) and scenario 2c (Regional stratified by port-quarter), as illustrated in Figure 3.c.

The second best-performing results are provided by scenarios 3a and 3c where the country stratification is extended to port and port-quarter stratifications, respectively (Figure 3.c). Regardless of the sampling design, the landings of both countries and both ICES Divisions are similarly estimated (Figures 3.d).

Scenarios 2a and 2c are the most precise overall but provide less coverage to some métiers and national fisheries (Table 3.c). Scenarios 3a and 3c have less precision in the estimation of total hake landings but improve the "tonnes of hake sampled" and "trips with hake sampled" while providing better coverage in some areas with smaller hake landings such as Southern Portugal (Table 3.d) and métiers such as PTB_MPD_>=55_0_0 (Figure 3.e).

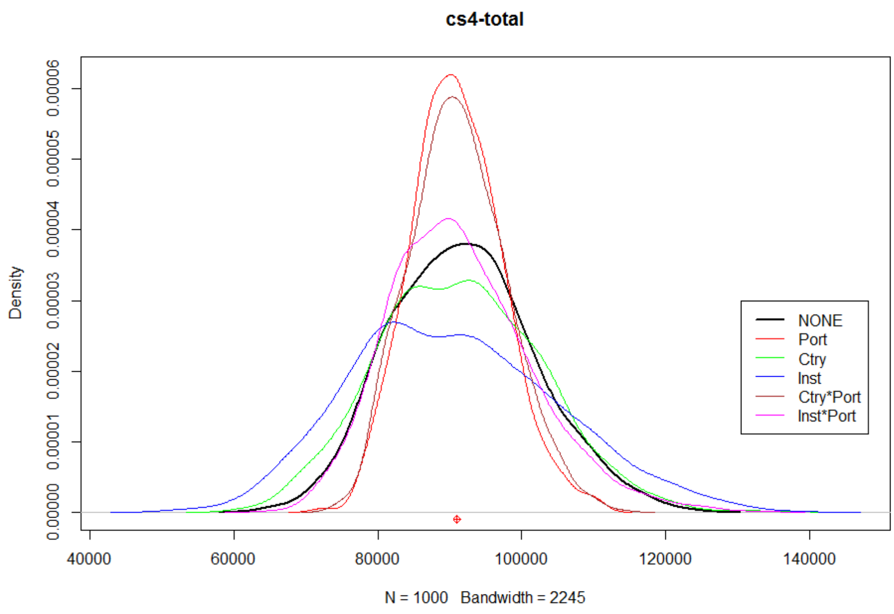


Figure 3.a. Density plot of estimated values (hake landings in tons) by scenario in CS4-shke. Red mark: true value.

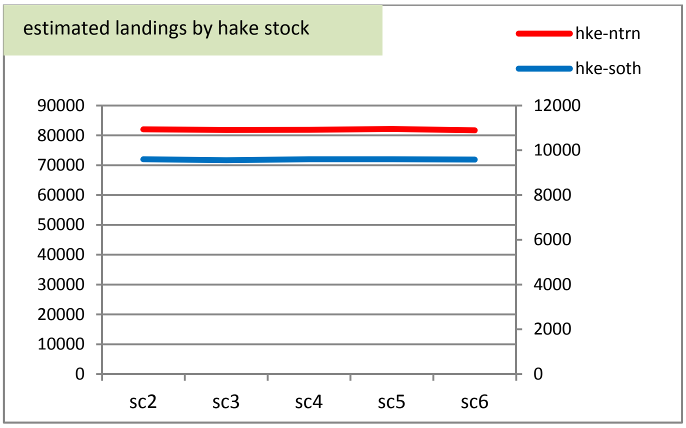


Figure 3.b. Estimated landings by CS4-total scenario by stock.

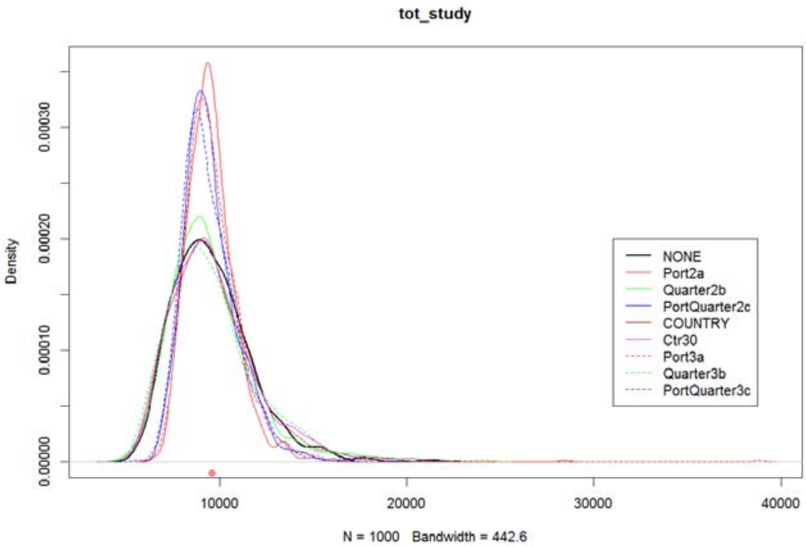


Figure 3.c. Density plot of estimated values (hake landings in tons) by scenario in CS4-shke. Red mark: true value. Note: Scenario Ctr30 = scenario 3d referred in text.

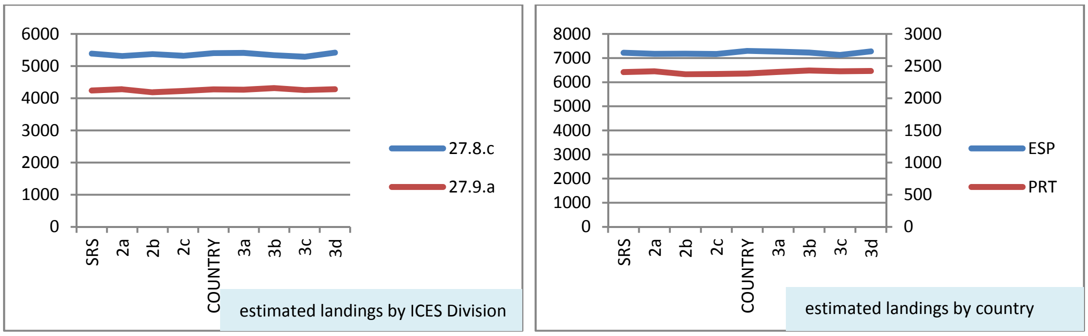


Figure 3.d. Estimated landings by CS4-shke scenario and ICES Division (left) and country (right).

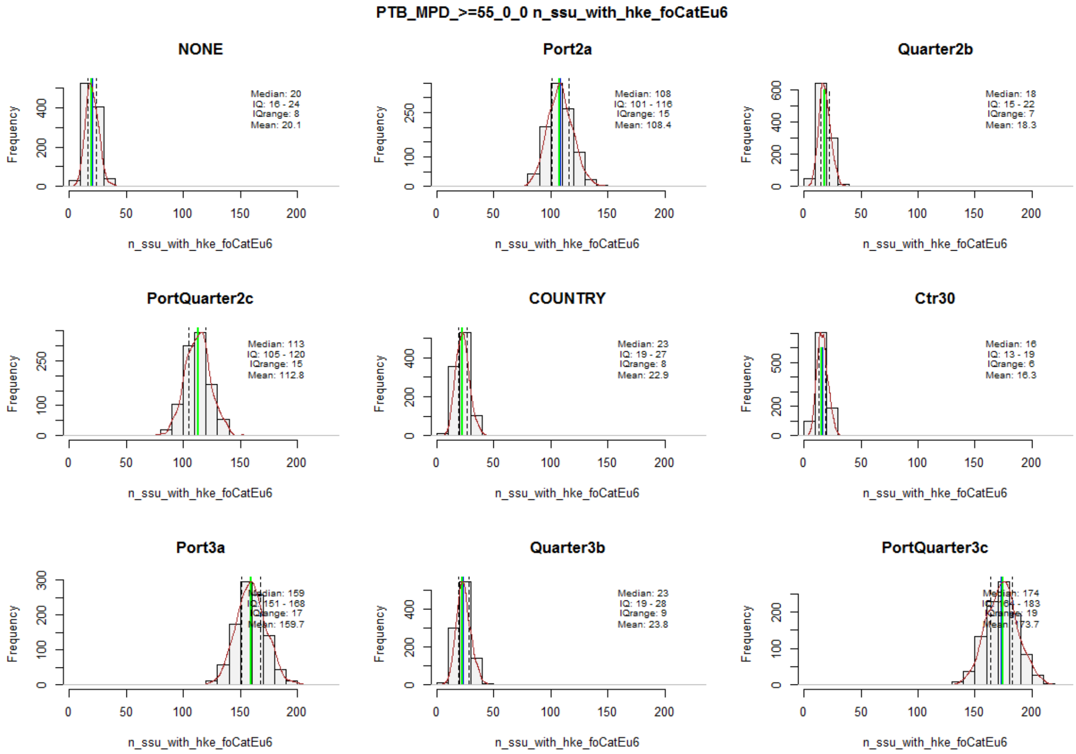


Figure 3.e. CS4-shke. Number of trips with hake sampled in métier PTB_MPD_>=55_0_0 under the different scenarios. Green line: median value. Dotted lines: interquartile range (25-75%). Brown line: density plot. Note: Scenario Ctr30 = scenario 3d referred in text.

Table 3.a. Results by scenario for CS4-total: bias is calculated as the ratio between estimated landings by scenario in relation to baseline scenario; precision is calculated as the ratio between SE of each scenario and the baseline scenario SE.

Scenario	Description of sampling design	Bias	Precision
1	2-stage SRS (no stratified)	baseline	baseline
2	Stratified by Country	1.00	1.14
3	Stratified by Institute	1.00	1.44
4	Regional stratified by Port	1.00	**0.63**
5	Stratified by Country and Port	1.00	**0.67**
6	Stratified by Institute and Port	1.01	1.00

Table 3.b. Results by scenario for CS4-shke: bias is calculated as the ratio between estimated landings by scenario in relation to baseline scenario; precision is calculated as the ratio between SE of each scenario and the baseline scenario SE.

Scenario	Description of sampling design	Bias	Precision
1	2-stage SRS WR (no stratified)	Baseline	Baseline
2a	Regional stratified by port	1.00	**0.67**
2b	Regional stratified by quarter	0.99	1.05
2c	Regional stratified by port*quarter	0.99	**0.68**
3	Stratified by country	1.01	1.11
3a	Stratified by country and port	1.00	0.80
3b	Stratified by country and quarter	1.00	1.17
3c	Stratified by country and port*quarter	0.99	0.76
3e	Stratified by country (proportional to landings)	1.01	1.09

Table 3.c. Average hake landings sampled and number of trips (SSU) with hake sampled by scenario and country for CS4-shke.

Parameter	Country	sc1	sc2a	sc2b	sc2c	sc3	sc3a	sc3b	sc3c	sc3d
Tons of hake sampled	ESP	53	182	50	187	54	**243**	53	**245**	45
	PRT	6	25	5	25	8	**33**	8	**32**	6
Number of trips with hake sampled	ESP	392	717	360	717	347	**732**	332	**725**	348
	PRT	98	244	89	241	135	**316**	129	**312**	99

Table 3.d. Average hake landings sampled and number of trips (SSU) with hake sampled by scenario and Iberian subarea for CS4-shke.

Parameter	Iberian subarea	sc1	sc2a	sc2b	sc2c	sc3	sc3a	sc3b	sc3c	sc3d
Tons of hake sampled	8cE1	11	13	10	16	8	12	8	13	10
	8cE2	20	61	20	66	30	153	31	157	15
	8cW	15	86	13	82	11	62	9	59	14
	9aC	1	7	1	9	1	5	1	59	1
	9aN	4	13	4	12	3	10	3	9	4
	9aS	2	3	2	2	1	2	1	1	2
	FOREIGN	1	2	1	2	0	1	0	1	1
	NW	2	9	2	9	3	13	3	12	2
	S	1	**0**	1	**0**	1	**0**	1	1	1
	SW	2	14	2	14	4	19	3	19	2
Number of trips with hake sampled	8cE1	104	70	100	98	83	53	81	76	95
	8cE2	52	97	49	107	93	281	99	285	34
	8cW	87	319	79	312	64	236	58	224	81
	9aC	6	22	6	18	4	15	4	12	5
	9aN	78	123	68	115	56	86	49	82	73
	9aS	64	87	58	66	47	61	41	46	60
	FOREIGN	7	20	7	19	6	16	5	14	7
	NW	41	100	36	95	58	134	56	128	42
	S	20	6	19	8	29	8	28	12	21
	SW	29	118	27	118	42	159	39	159	30

4 Conclusions

The results of simulating different sampling designs were evaluated to select the most robust sampling plan from a statistical point of view. For this, a protocol was established in order to follow a selection process as objective as possible.

4.1 Protocol of selection

The protocol of selection consists of the following main steps:

1. Bias analyses: all scenarios should be unbiased or have their biases quantified. All analyses carried out in CS4 were unbiased (Tables 3.a and 3.b).
2. Precision analyses: selection of scenarios with highest precision (Table 3.a and 3.b for CS4-total and CS4-shke, respectively). Precision may be evaluated as standard error (particularly useful with Gaussian distributions) and/or inter-quartile range (more adequate with skewed distributions).
3. Coverage by country: regardless of the statistical robustness it is necessary to analyze the possible deviations that may occur with respect to the current coverage of the National sampling programs. This is an aspect that may be relevant when discussing a feasibility analysis.
4. Coverage by domain: regardless of the statistical robustness it is necessary to analyze the possible deviations that may occur with respect to the post-stratification currently used by the National institutes to provide commercial data to ICES: technical (métier), spatial (ICES Division), and time (quarter) disaggregation. This is an aspect that may be relevant when discussing a feasibility analysis. Evidence from CS4 results shows that higher coverage of some areas and métiers may be achieved by some scenarios that yielded slightly lower (but still acceptable) precision in estimates of total landings.
5. Realism: all scenarios should match real-life sampling design conditions. As an example, all days of the year should be included unless specific days are known not to have any landings; all fleets should be included unless specific fleets are known to have negligible landings with similar fish size/age structures. In this study CS4-shake considers all fishing activity and weekdays in their simulations, complying with the agreed fishPi WP2 approach. Results from CS4-total are based only on days with hake landings but should also be considered for the feasibility study as CS4-shke covers the Southern hake stock, which represents only 10% of the European hake landings.
6. Adequacy to end-user needs: A regional sampling design should aim at estimating accurate length/age composition of the landings of the hake stocks for stock assessment purposes while at the same time providing accurate estimates for many other stocks and fisheries subject to assessment. The current study focused on estimating total landings of hake which are hardly a proxy for size/age structure (e.g., due to natural and fishing mortality landings of large/older fish are frequently a small component of total landings) and which main fleets may not encompass the main fleets of other exploited stocks.

5 Recommendation

The first two points of the protocol of selection indicate higher precision for regional scenarios with port stratification. This approach was implemented differently depending on the data set. CS4-total was stratified into two port strata, one grouping the main ports and other the secondary ports. However CS4-shke was stratified in a higher number of ports, so the most important ports are used as individualized strata.

When the results are analyzed in more detail it is evident that the countries or areas with low catches loose representation in regional scenarios. A feasibility study should address these issues as European sampling programs, in addition to serving the CFP objectives, are also used to provide local advice, from national governments to the fisheries sector partnerships. Moreover, the feasibility study should also address the existing practical constraints and how they may affect the theoretical sampling designs.

Appendix 1 Cs4- Total simulation scripts

```
##################################################
###       This is the R code for CS4 hke-total simulations       ###
###                Simulations were run using R 3.2.2            ###
##################################################
rm(list=(ls()))
library(data.table)
library(sampling)
library(survey)
library(lattice)
library(latticeExtra)
library (stringr)
library(dplyr)
# set wd
 setwd("...")
 source("sample from a group 2.R")
# read population data
 load("data")
 df1<-cs42014
# onShoreSampCtr different than c("ES","FR","IE","GB","DK","PT","NL","SE") are removed
 df1 <- subset(df1, onShoreSampCtr!="OT")

###########################################################
#     set sampling effort (magic numbers)
###########################################################
 # sampling effort by Sampling Institute
 DK <- 113;  ES_AZTI <- 43 ; ES_IEO <- 765;   FR <- 597;   GB_CEFAS <- 431;   GB_MS <- 45;
 IE <- 82;   NL <- 38;      PT <- 700;      SE <- 3
 # sampling effort by Country
 ES <- ES_IEO + ES_AZTI;
 GB <- GB_CEFAS + GB_MS
 ntotsamp <- sum(c(DK, ES, FR, GB, IE, NL, PT, SE) )   # 2817
 n_trips_per_market_event<-1

###########################################################
#     Define strata.
###########################################################
# From the following lines, select the strata you want to use.
# Sampling effort will need to be set accordingly (below)
##df1$stratum <- "ALL";  strat <- "ALL"  # scenario without stratification (two-stage random sampling): baseline
#df1$stratum <- df1$onShoreSampCtr;  strat <- "SampCtr"
#df1$stratum <- df1$portStrata;  strat <- "Port"
#df1$stratum <- df1$quarter;  strat <- "Q"
#df1$stratum <- paste(df1$onShoreSampCtr,df1$quarter, sep=".");  strat <- "SampCtrQ"
#df1$stratum <- paste(df1$onShoreSampCtr,df1$portStrata, sep=".");  strat <- "SampCtrPort"
#df1$stratum <- df1$onShoreSampIns;  strat <- "SampIns"
#df1$stratum <- paste(df1$onShoreSampIns,df1$quarter, sep=".");  strat <- "SampInsQ"
#df1$stratum <- paste(df1$onShoreSampIns,df1$portStrata, sep=".");  strat <- "SampInsPort"
#Define the sampling effort per strata
# without stratification (strat = "ALL")
if(strat=="ALL") {
 sampef <- c("ALL" = ntotsamp)
}
# stratification by country (strat = "SampCtr")
if(strat=="SampCtr") {
 sampef <- c("DK"=DK,"ES"=ES, "FR"=FR,"GB"=GB,"IE"=IE, "NL"=NL,"PT"=PT,"SE"=SE)
}
# stratification by port (strat = "Port")
if(strat=="Port") {
 port <-round(prop.table(tapply(df1$landWt, df1$stratum, sum))*ntotsamp,0)
 sampef <- c(port["major"], port["minor"])
}
# stratified by quarter (strat = "Q")
if(strat=="Q") {
 Q <- round(ntotsamp/4, digits=0)
 sampef <- c("Q1"=Q,"Q2"=Q, "Q3"=Q,"Q4"=Q)
}
```

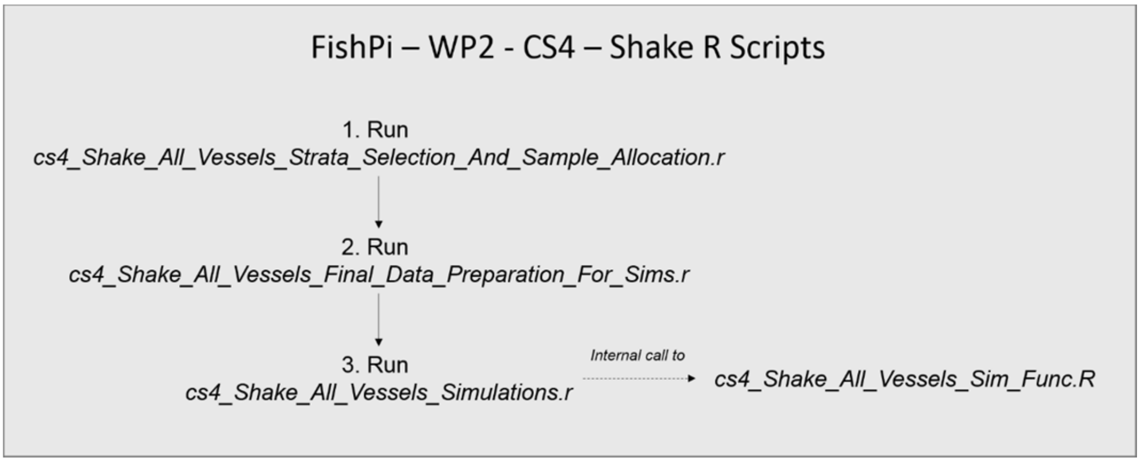
```r
# stratified by country and quarter (strat="SampCtrQ")
if(strat=="SampCtrQ") {
 DKq <- round(DK/4, digits=0);  ESq <- round(ES/4, digits=0);  FRq <- round(FR/4, digits=0);  GBq <- round(GB/4, digits=0);
 IEq <- round(IE/4, digits=0);  NLq <- round(NL/4, digits=0);  PTq <- round(PT/4, digits=0);  SEq <- round(SE/4, digits=0)
  sampef <- c("DK.Q1"=DKq, "DK.Q2"=DKq, "DK.Q3"=DKq, "DK.Q4"=DKq,
       "ES.Q1"=ESq, "ES.Q2"=ESq, "ES.Q3"=ESq, "ES.Q4"=ESq,
       "FR.Q1"=FRq, "FR.Q2"=FRq, "FR.Q3"=FRq,  "FR.Q4"=FRq,
       "GB.Q1"=GBq,  "GB.Q2"=GBq, "GB.Q3"=GBq, "GB.Q4"=GBq,
       "IE.Q1"=IEq, "IE.Q2"=IEq, "IE.Q3"=IEq, "IE.Q4"=IEq,
       "NL.Q1"=9, "NL.Q2"=NLq, "NL.Q3"=NLq, "NL.Q4"=NLq,
       "PT.Q1"=PTq, "PT.Q2"=PTq, "PT.Q3"=PTq, "PT.Q4"=PTq,
       "SE.Q1"=SEq, "SE.Q2"=SEq, "SE.Q3"=SEq, "SE.Q4"=SEq)
}
# stratified by country and port (strat="SampCtrPort")
if(strat=="SampCtrPort") {
 portCtr <-tapply(df1$landWt, list(df1$onShoreSampCtr,df1$portStrata), sum, na.rm=TRUE )
 portCtr[which(is.na(portCtr))]<-0
 portCtr <- data.frame(prop.table(portCtr,1) )
 portCtr$samp <- c(DK,ES,FR,GB,IE,NL,PT,SE)
 portCtr$samp.major <-   round(portCtr$major * portCtr$samp,0)
 portCtr$samp.minor <-   round(portCtr$minor * portCtr$samp,0)
 sampef <- c("DK.major"=portCtr["DK","samp.major"], "DK.minor"=portCtr["DK","samp.minor"],
       "ES.major"=portCtr["ES","samp.major"], "ES.minor"=portCtr["ES","samp.minor"],
       "FR.major"=portCtr["FR","samp.major"], "FR.minor"=portCtr["FR","samp.minor"],
       "GB.major"=portCtr["GB","samp.major"], "GB.minor"=portCtr["GB","samp.minor"],
       "IE.major"=portCtr["IE","samp.major"], "IE.minor"=portCtr["IE","samp.minor"],
       "NL.minor"=portCtr["NL","samp.minor"],
       "PT.minor"=portCtr["PT","samp.minor"],
       "SE.minor"=portCtr["SE","samp.minor"])
}
# stratified by sampling institute (strat="SampIns")
if(strat=="SampIns") {
sampef    <-    c("DK"=DK,"ES_AZTI"=ES_AZTI,     "ES_IEO"=ES_IEO,"FR"=FR,"GB_CEFAS"=GB_CEFAS,"GB_MS"=GB_MS,     "IE"=IE,
"NL"=NL,"PT"=PT,"SE"=SE)
}
# stratified by sampling institute and quarter (strat="SampInsQ")
if(strat=="SampInsQ") {
 DKq <- round(DK/4, digits=0);   ES_AZTIq <- round(ES_AZTI/4, digits=0);     ES_IEOq <- round(ES_IEO/4, digits=0);
 FRq <- round(FR/4, digits=0);   GB_CEFASq <- round(GB_CEFAS/4, digits=0);    GB_MSq <- round(GB_MS/4, digits=0);
 IEq <- round(IE/4, digits=0);   NLq <- round(NL/4, digits=0);             PTq <- round(PT/4, digits=0);
 SEq <- round(SE/4, digits=0)
 sampef <- c("DK.Q1"=DKq, "DK.Q2"=DKq, "DK.Q3"=DKq, "DK.Q4"=DKq,
      "ES_AZTI.Q1"=ES_AZTIq, "ES_AZTI.Q2"=ES_AZTIq, "ES_AZTI.Q3"=ES_AZTIq, "ES_AZTI.Q4"=ES_AZTIq,
      "ES_IEO.Q1"=ES_IEOq, "ES_IEO.Q2"=ES_IEOq, "ES_IEO.Q3"=ES_IEOq, "ES_IEO.Q4"=ES_IEOq,
      "FR.Q1"=FRq, "FR.Q2"=FRq, "FR.Q3"=FRq,  "FR.Q4"=FRq,
      "GB_CEFAS.Q1"=GB_CEFASq,  "GB_CEFAS.Q2"=GB_CEFASq, "GB_CEFAS.Q3"=GB_CEFASq, "GB_CEFAS.Q4"=GB_CEFASq,
      "GB_MS.Q1"=GB_MSq,  "GB_MS.Q2"=GB_MSq, "GB_MS.Q3"=GB_MSq, "GB_MS.Q4"=GB_MSq,
      "IE.Q1"=IEq, "IE.Q2"=IEq, "IE.Q3"=IEq, "IE.Q4"=IEq, "NL.Q1"=9, "NL.Q2"=NLq, "NL.Q3"=NLq, "NL.Q4"=NLq,
      "PT.Q1"=PTq, "PT.Q2"=PTq, "PT.Q3"=PTq, "PT.Q4"=PTq, "SE.Q1"=SEq, "SE.Q2"=SEq, "SE.Q3"=SEq, "SE.Q4"=SEq)
}
# stratified by sampling institute and port  (strat="SampInsPort")
if(strat=="SampInsPort") {
 portIns <-tapply(df1$landWt, list(df1$onShoreSampIns,df1$portStrata), sum, na.rm=TRUE )
 portIns[which(is.na(portIns))]<-0
 portIns <- data.frame(prop.table(portIns,1) )
 portIns$samp <- c(DK,ES_AZTI, ES_IEO,FR,GB_CEFAS, GB_MS,IE,NL,PT,SE)
 portIns$samp.major <-   round(portIns$major * portIns$samp,0)
 portIns$samp.minor <-   round(portIns$minor * portIns$samp,0)
 sampef <- c("DK.major"=portIns["DK","samp.major"],           "DK.minor"=portIns["DK","samp.minor"],
       "ES_AZTI.major"=portIns["ES_AZTI","samp.major"],   "ES_AZTI.minor"=portIns["ES_AZTI","samp.minor"],
       "ES_IEO.major"=portIns["ES_IEO","samp.major"],    "ES_IEO.minor"=portIns["ES_IEO","samp.minor"],
       "FR.major"=portIns["FR","samp.major"],           "FR.minor"=portIns["FR","samp.minor"],
       "GB_CEFAS.minor"=portIns["GB_CEFAS","samp.minor"],
       "GB_MS.major"=portIns["GB_MS","samp.major"],     "GB_MS.minor"=portIns["GB_MS","samp.minor"],
       "IE.major"=portIns["IE","samp.major"],          "IE.minor"=portIns["IE","samp.minor"],
       "NL.minor"=portIns["NL","samp.minor"],
       "PT.minor"=portIns["PT","samp.minor"],
       "SE.minor"=portIns["SE","samp.minor"])
}
(sum(sampef))   # 2817
```

```
####################################################################
#   Prepare the data
####################################################################
df1$landWt<- df1$landWt/1000
df1<-df1[order(stratum, onShoreSampLoc, arvDate)]
# simplify the matrix
 df1<-df1[,list(landWt=sum(landWt)), by=.(stratum=stratum, market=onShoreSampLoc, market_day=market_day,
                          market_day_event=market_day_event, arvDate=arvDate, voyageId=voyageId, vslId=vslId,
                          domainNat1=domainNat1, domainNat2=domainNat2)]
# total population size in each stratum
 table(df1$stratum)
 df1<- df1 [,"Ntrips_total":=length(unique(market_day_event)), ] []                    # total n of trips
 df1<- df1[,"Ntrips_strata":=length(unique(market_day_event)), by=stratum] []              # n of trips per strata
 df1<- df1 [,"Ntrips_marketday":=length(unique(market_day_event)), by=.(stratum,market_day)] [] # n of trips per market day
 df1<- df1[,"Nmarketdays_strata":=length(unique(market_day)), by=stratum] []               # n of market days per strata
 df1<- df1 [,"Nmarketdays_strata_sampled":=sampef[match(df1$stratum, names(sampef))],] []      # n of market days sampled per
strata
 # summary by strata.
sum_strat <- df1[,list("landWt"=sum(landWt),
                "Nmarketdays_strata"=length(unique(market_day)),
                "Ntrips_strata"=length(unique(market_day_event))),
            by=stratum]
sum_strat$Nsampled_marketdays_strata <- sampef
write.csv(sum_strat, file=paste("SampEffort_", strat,".csv" sep=""), row.names=FALSE)
# We define pop as the data we will use in the simulations
pop<-df1

####################################################################################
##   This is a function to select the samples according to a stratified two-stage sampling design
####################################################################################
# Strat defines the strata (ie landing country). If there is no stratification it can e set to a constant
# Clus1 defines the PSU to be sampled  (ie market day)
# Clus2 defines the SSU to be sampled  (ie market day event or trip)
# SizeClus1 is a vector giving the sample sizes in strata
# SizeClus2 is a constant giving the sample sizes in psu
# the values in Strat need to match the names of SizeClus1
sampleFromAGroup2 <- function(Strat,Clus1,Clus2,SizeClus1,SizeClus2) {
           #package needed:
           #Â dplyr   : sample_n function to sample n individuals according to a group
           # sampling: strata function to sample n1,n2,n3... individuals according to n groups
           if(!require(dplyr)){
                     print("dplyr package is missing")
           }else{library(dplyr)}
           if(!require(sampling)){
                     print("sampling package is missing")
           }else{library(sampling)}
           require(sampling)
           ################################################################
           #1. select  n Clus1 taken randomly by Strat without replacement, n given by SizeClus1
           #1.1 format SizeClus1 in data.frame (order of Strat between Strat and SizeClus1 has to be the same)
           sampef2<- data.frame(Strat=names(sampef),n=sampef)%>%arrange(Strat)
           aa<-tbl_df(data.frame(Strat=Strat,Clus1=Clus1,Clus2=Clus2))%>%arrange(Strat)
           #1.2 keep only Strat and Clus1
           aa1<-distinct(aa[,1:2])
           #1.3 sample n Clus1 by Strat according to SizeClus1 without replacement (srswor method)
           pipo<-strata(aa1,"Strat",sampef2$n,meth="srswor")
           #1.4 get the data
           aa1<-getdata(aa1,pipo)
           ################################################################
           #2. select n Clus2 in the previously selected Clus1 with replacement with n=SizeClus2
           #2.1 filter Clus2: keep only Clus2 who are in Clus1 selected previously
           aa2<-aa%>%filter(Clus1%in%aa1$Clus1)%>%select(Clus1,Clus2)%>%distinct()
           #2.2 sample n Clus2 in Clus1 with n=SizeClus2
           pipo<-aa2%>%group_by(Clus1)%>%sample_n(size=SizeClus2,replace=F)%>%ungroup()
           #2.3 return Clus2
           return(pipo$Clus2)
 }

####################################################################
#   Simulations
####################################################################
```

```
Nrep      <- 1500      ##---number of replicates
# creates de output matrix (estimations of totals)
ressrs   <-   data.frame(simID   =   rep(seq(1,Nrep),each=1),   domainNat=rep(1,each=1,   times=Nrep),   esttotal=rep(NA,Nrep*1), CIlow=rep(NA,Nrep*1), CIhigh=rep(NA,Nrep*1))
# creates de output matrix (poststratification domain 1)
level_domain1 <- length(unique((pop$domainNat1)))
domain_list1 <- unique(pop$domainNat1)
ressrs_p1   <-   data.frame(simID   =   rep(seq(1,Nrep),each=level_domain1),   domainNat=rep(unique((pop$domainNat1)),each=1, times=Nrep), esttotal=rep(NA,Nrep*level_domain1), CIlow=rep(NA,Nrep*level_domain1), CIhigh=rep(NA,Nrep*level_domain1))
ressrs_p1$ID <- paste(ressrs_p1$simID, ressrs_p1$domainNat)
# creates de output matrix (poststratification domain 2)
level_domain2 <- length(unique((pop$domainNat2)))
domain_list2 <- unique(pop$domainNat2)
ressrs_p2   <-   data.frame(simID   =   rep(seq(1,Nrep),each=level_domain2),   domainNat=rep(unique((pop$domainNat2)),each=1, times=Nrep), esttotal=rep(NA,Nrep*level_domain2), CIlow=rep(NA,Nrep*level_domain2), CIhigh=rep(NA,Nrep*level_domain2))
ressrs_p2$ID <- paste(ressrs_p2$simID, ressrs_p2$domainNat)

for (i in 1:Nrep) {
  # step1: select the sample
 selected   <-   sampleFromAGroup2   (Strat=pop$stratum,   Clus1=pop$market_day,   Clus2=pop$market_day_event,   SizeClus1= sampef, SizeClus2=n_trips_per_market_event )
 samp <- pop[pop$market_day_event %in% selected,]

# step2: define probabilities and describe the sampling design
 samp$Prob1<- samp$Nmarketdays_strata_sampled/samp$Nmarketdays_strata              # prob cluster 1
 samp$Prob2<- n_trips_per_market_event/samp$Ntrips_marketday                      # prob cluster 2
 samp$Prob <- samp$Prob1 * samp$Prob2
 samp$fpc1 <- samp$Ntrips_strata                                    # total population in each stratum
 samp$fpc2 <- samp$Ntrips_total
  #dSRS<-svydesign(id=~market_day+market_day_event, probs=~Prob, strata=~stratum, data=samp, nest=TRUE, fpc=~fpc1+fpc2)
 dSRS<-svydesign(id=~market_day+market_day_event, probs=~Prob, strata=~stratum, data=samp, nest=TRUE)

 ##---step3: estimate the total for a given sampling design---
 options(survey.lonely.psu = "adjust")
 out <- svytotal(~landWt, dSRS, na.rm=FALSE, deff=T)
 out <- data.frame(simID=i, landWt=out[1], CIlow=as.numeric(out[1]-1.96*SE(out)), CIhigh=as.numeric(out[1]+1.96*SE(out)))
 out$ID <- out$simID
 ressrs[match(out$simID, ressrs$simID),c(3:5)] <- out[,c("landWt", "CIlow", "CIhigh")]
 print(paste(i,"  total estimation"))
 ##---step4.1: Domain1  -  estimate the total by poststrata for a given sampling design---
 poststrat.weights <- data.frame(table(domain=pop$domainNat1))
 names(poststrat.weights) <- c("domainNat1","Freq")
 ##---------- Here we need to put partial=TRUE, because some of domains may be missing in the object dSRS
 ##           Then we ignore these messing domains in post-stratification----------##
 dSRSpoststr <- postStratify(dSRS, strata =~domainNat1, population = poststrat.weights, partial = TRUE)
 svytotal(~landWt, dSRSpoststr, na.rm=F, deff=T)
 outpost <- svyby(~landWt, ~domainNat1, dSRSpoststr, svytotal)
 outpost <- data.frame(outpost, simID=i, CIlow=outpost$landWt-1.96*outpost$se, CIhigh=outpost$landWt+1.96*outpost$se)
 outpost$ID <- paste(outpost$simID, outpost$domainNat1)
 ressrs_p1[match(outpost$ID, ressrs_p1$ID),c(3:5)] <- outpost[,c("landWt", "CIlow", "CIhigh")]
 print(paste(i,"  poststratification"))
 ##---step4.2: estimate the total by poststrata for a given sampling design--- DOMAIN 2
 poststrat.weights <- data.frame(table(domain=pop$domainNat2))
 names(poststrat.weights) <- c("domainNat2","Freq")
 ##---------- Here we need to put partial=TRUE, because some of domains may be missing in the object dSRS
 ##           Then we ignore these messing domains in post-stratification----------##
 dSRSpoststr <- postStratify(dSRS, strata =~domainNat2, population = poststrat.weights, partial = TRUE)
 svytotal(~landWt, dSRSpoststr, na.rm=F, deff=T)
 outpost <- svyby(~landWt, ~domainNat2, dSRSpoststr, svytotal)
 outpost <- data.frame(outpost, simID=i, CIlow=outpost$landWt-1.96*outpost$se, CIhigh=outpost$landWt+1.96*outpost$se)
 outpost$ID <- paste(outpost$simID, outpost$domainNat2)
 ressrs_p2[match(outpost$ID, ressrs_p2$ID),c(3:5)] <- outpost[,c("landWt", "CIlow", "CIhigh")]
 print(paste(i,"  poststratification"))
}

# for the estimation of totals
res_pop <- sum(pop$landWt)    ##-----  population total
est_pop <- mean(ressrs$esttotal)    ##-----  estimated total
save(ressrs,               res_pop,               est_pop,              file=paste("cs4_StratTwoStag_",strat,"_sim_totals_",format(Sys.time(), "%Y%m%d%H%M"),".RData",sep=""))
```

```
# for the postratification --  DOMAIN1
res_pop_p1 <- aggregate(landWt~domainNat1, FUN=sum, data=pop)    ##-----  population total per domain
est_pop_p1 <- aggregate(esttotal~domainNat, FUN=mean, data=ressrs_p1, na.rm=TRUE, na.action=NULL)    ##-----  estimated total per domain
est_pop_p1$npsu<- tapply(pop$market_day_event, list(pop$domainNat1), function(x){length(unique(x))})
est_pop_p1$nsim <- tapply(ressrs_p1$esttotal, list(ressrs_p1$domainNat), function(x){length(x[which(!is.na(x))])})
sum(est_pop_p1$esttotal, na.rm=TRUE)
save(ressrs_p1,       res_pop_p1,       est_pop_p1,       file=paste("cs4_StratTwoStag_",strat,"_sim_poststrat1_",format(Sys.time(), "%Y%m%d%H%M"),".RData",sep=""))

# for the postratification --  DOMAIN2
res_pop_p2 <- aggregate(landWt~domainNat2, FUN=sum, data=pop)    ##-----  population total per domain
est_pop_p2 <- aggregate(esttotal~domainNat, FUN=mean, data=ressrs_p2, na.rm=TRUE, na.action=NULL)    ##-----  estimated total per domain
est_pop_p2$npsu<- tapply(pop$market_day_event, list(pop$domainNat2), function(x){length(unique(x))})
est_pop_p2$nsim <- tapply(ressrs_p2$esttotal, list(ressrs_p2$domainNat), function(x){length(x[which(!is.na(x))])})
sum(est_pop_p1$esttotal, na.rm=TRUE)
save(ressrs_p2,       res_pop_p2,       est_pop_p2,       file=paste("cs4_StratTwoStag_",strat,"_sim_poststrat2_",format(Sys.time(), "%Y%m%d%H%M"),".RData",sep=""))
```

Appendix 2 CS4- shke simulation scripts




```
# cs4_Shake_All_Vessels_Strata_Selection_And_Sample_Allocation.r
# ================================================
# fishPi - WP2 - CS2 - SHake - Strata Selection And Sample Allocation - 'All Vessels' Setting
# ================================================

 # note: files overwrite on each run
 # note: csv files contains the number of samples per strata to be used in simulations
 # note: updates fishPi object with strata information


# ====================================
# Packages and data
# ====================================

 # load packages
  library(stringr)
  library(reshape)

 # Reads data
  db1<-readRDS("cs4_all_checked_Shake_correct.rds")

# ====================================
# Creation of onShoreSampCtr (PT, ES, BC OT)
# ====================================

 # creates onShoreSampCtr from OnshoreSampLoc
  BCPorts<- c("ESALT", "ESARW", "ESBID", "ESBIO", "ESBRM", "ESCIV", "ESDAC", "ESDRI", "ESDVA",
     "ESEAS", "ESEEM", "ESEWE", "ESGTI", "ESGET", "ESIRU", "ESLEK", "ESLEQ", "ESMTU",
     "ESOND", "ESORI", "ESPAS", "ESPOR", "ESRTA", "ESSNR", "ESVIT", "ESZAU", "ESZOD")

  str_sub(db1$onShoreSampLoc[which(db1$onShoreSampLoc %in% BCPorts)],1,2) <- "BC"
  db1$onShoreSampCtr<-str_sub(db1$onShoreSampLoc,1,2)

   # creates onShoreSampCtr = OT = OTHER
    db1$onShoreSampCtr[!db1$onShoreSampCtr %in% c("PT","ES", "BC")]<-"OT"

 # creates onShoreSampQuarter from arvDate
  db1$onShoreSampQuarter<-quarters(db1$arvDate)

 # subsets 2013 [design dataset]
  df1<-droplevels(db1[format(db1$arvDate, "%Y" )=="2013",])

# ====================================
# Magic Numbers
# ====================================

 # defines magic numbers
```

```
 #PRT: 872 market*days; 1409 trips (trips per day: 1.6)
 #ESP: 1425 market*days (+48+62 from Cadiz at-sea); 2382 trips (trips per day: 1.55 with Cadiz)
 #TOTAL: 2297 market*days (+48+62 from Cadiz at-sea); 3791 trips (trips per day: 1.57 with Cadiz)

 psuType="market*days" # note: psuType is currently only used for axis labels

 psuPT<-872
 psuES<-1535-391
 psuBC<-391
 psuOT<-0
 psuALL<-psuPT+psuES+psuBC+psuOT

# =====================================
# Allocaton settings
# =====================================

 # sets minimum number of market_days (psus) and trips (ssus)
 min_psu_per_strata<-2
 n_ssu_per_psu<-2

 # type of allocation
 #allocType="n_strata" # selects n strata regardless of the x% of landings they represent (=> n strata fixed)
 allocType="perc_land" # selects strata up to x% of landings (=> n strata variable)

 # set maximum number of strata if allocType=="n_strata"
 if(allocType == "n_strata")
  {
  n_main_strata_regional<-19
  n_main_strata_national_pt<-9
  n_main_strata_national_es<-9
  n_main_strata_national_bc<-9
  }

 # set % land for stratification if allocType=="perc_land"
 if (allocType == "perc_land")
  {
  perc_land_regional<-75
  perc_land_regional_pt<-75
  perc_land_regional_es<-75
  perc_land_regional_bc<-75
  }

# =====================================
# production of sample distributions (files samp_dist)
# =====================================

 # =====================================
 # Scenario 1 [random sampling]
 # =====================================

  res<-data.frame(NONE = "NONE", n_psu= psuALL, n_ssu_per_psu=n_ssu_per_psu)
  write.csv2(res,"Results\\Shake_AllVessels_ScN_samples.csv", row.names=FALSE)


 # =====================================
 # Scenario 2 [regional]
 # =====================================

  # =====================================
  # Port strata 2013 (2a)
  # =====================================


  weight.strata<-sort(tapply(df1$landWt, df1$onShoreSampLoc, sum), decreasing=TRUE)

  # raw plots
   windows(15,7)
   par(mfrow=c(1,2))
   barplot(weight.strata/1000, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
   plot(ecdf(weight.strata/1000))
```

```
   weight.strata[1:25]/1000
   round(cumsum(weight.strata)/sum(weight.strata)*100,0) [1:25]

   dev.off()

  # selects de n_main_strata_regional portos + OTHER #
   if(allocType == "perc_land")
    {
    n_main_strata_regional<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional))+1
    }
   percland.main.strata<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_regional]
   percland.main.strata

   df1$Port2a<-NA
   df1$Port2a[df1$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_regional]]<-
df1$onShoreSampLoc[df1$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_regional]]
   df1$Port2a[!df1$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_regional]]<-"OT999"

  # barplot (landings)
   windows(15,7)
   par(mfrow=c(1,2))
   weight.strata<-sort(tapply(df1$landWt, df1$Port2a, sum)/1000, decreasing=TRUE)
   cores<-str_sub(names(weight.strata),1,2)
   cores[cores=="PT"]<-"red"
   cores[cores=="ES"]<-"blue"
   cores[cores=="BC"]<-"green"
   cores[cores=="OT"]<-"gray"
   barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
   a<-tapply(df1$landWt, str_sub(df1$Port2a,1,2), sum)
   barplot(a/1000, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = "Landings
(ton)", xlab = "Institute")
   title(main="SHake (All Vessels Setting) - Scenario 2a", outer=TRUE, line=-1)
   title(main=paste("main strata (",length(weight.strata[!grepl(names(weight.strata), pat="999")]),") =",percland.main.strata,"% kg; all
strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
   savePlot(file="Results\\Shake_AllVessels_Sc2a_landings.png",type="png")

  # sample allocation
   samples<-round(prop.table(tapply(df1$landWt, df1$Port2a, sum))*psuALL,0) # 3792

  # barplots (psus to sample)
   windows(15,7)
   par(mfrow=c(1,2))
   cores<-str_sub(names(samples),1,2)
   cores[cores=="PT"]<-"red"
   cores[cores=="ES"]<-"blue"
   cores[cores=="BC"]<-"green"
   cores[grepl(names(samples), pat="999")]<-"gray"
   barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports")
   a<-tapply(samples, str_sub(names(samples),1,2), sum)
   barplot(a, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = psuType, xlab =
"Institute")
   title(main="Sample Distribution SHake (All Vessels Setting)  - Scenario 2a", outer=TRUE, line=-1)
   title(main=paste("(strata <2 samples = ", sum(samples<1),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
pat="999")]),") =",percland.main.strata,"% kg; all strata = 100% kg)"), outer=TRUE, line=-2, cex.main=0.9)
   savePlot(file="Results\\Shake_AllVessels_Sc2a_samples.png",type="png")

  # csv ouput
   res<-melt(samples)
   colnames(res)[1]<-"Port2a"
   colnames(res)[2]<-"n_psu"
   res$n_ssu_per_psu=n_ssu_per_psu
   res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata; sum(res[,2]); sum(res[,2]*res[,3])
   write.csv2(res,"Results\\Shake_AllVessels_Sc2a_samples.csv", row.names=FALSE)

  graphics.off()

 # =====================================
 # Quarter strata 2013 (2b)
 # =====================================

  weight.strata<-sort(tapply(df1$landWt, df1$onShoreSampQuarter, sum), decreasing=TRUE)
```

```
# raw plots
 windows(15,7)
 par(mfrow=c(1,2))
 barplot(weight.strata, las=2, cex.names=0.5)
 plot(ecdf(weight.strata/1000))

 dev.off()

 round(cumsum(weight.strata)/sum(weight.strata)*100,0)

 # note: an option was made for including all quarters since 75% land generally implies one quarter is left out
 if(allocType == "perc_land")
  {
  #n_main_strata_regional<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional))+1
  n_main_strata_regional<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<=100))
  }
 percland.main.strata<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_regional]
 percland.main.strata

 df1$Quarter2b<-df1$onShoreSampQuarter

# barplot (landings)
 par(mfrow=c(1,1))
 barplot(weight.strata/1000, las=2, cex.names=0.5)
 title(main=c("SHake (All Vessels Setting) - Scenario 2b"), outer=TRUE, line=-1)
  title(main=paste("(main strata (",length(weight.strata[!grepl(names(weight.strata), pat="999")]),") =",percland.main.strata,"% kg; all
strata = 100% kg)"), outer=TRUE, line=-2, cex.main=0.9)
  savePlot(file="Results\\Shake_AllVessels_Sc2b_landings.png",type="png")

# sample allocation
 samples<-round(prop.table(tapply(df1$landWt, df1$Quarter2b, sum))*psuALL,0) # 3791

# barplots (psu)
 par(mfrow=c(1,1))
 cores<-str_sub(names(samples),1,2)
 cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
 cores[grepl(cores, pat="Q1")]<-"red"
 cores[grepl(cores, pat="Q2")]<-"blue"
 cores[grepl(cores, pat="Q3")]<-"green"
 cores[grepl(cores, pat="Q4")]<-"pink"
 cores[grepl(names(samples), pat="999")]<-"gray"
 barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Quarters")
 title(main="Sample Distribution SHake (All Vessels Setting) - Scenario 2b", outer=TRUE, line=-1)
  title(main=paste("(strata <2 samples = ", sum(samples<2),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
pat="999")]),") =",percland.main.strata,"% kg; all strata = 100% kg)"), outer=TRUE, line=-2, cex.main=0.9)
  savePlot(file="Results\\Shake_AllVessels_Sc2b_samples.png",type="png")

# csv ouput
 res<-melt(samples)
 colnames(res)[1]<-"Quarter2b"
 colnames(res)[2]<-"n_psu"
 res$n_ssu_per_psu=n_ssu_per_psu
 res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata; sum(res[,2]); sum(res[,2]*res[,3])
 write.csv2(res,"Results\\Shake_AllVessels_Sc2b_samples.csv", row.names=FALSE)

graphics.off()

# ====================================
# Port vs Quarter strata 2013 (2c)
# ====================================

 # Note: Two ways could be (and were!) considered: a) use the 'major' ports previously selected and b) use all port*quarter
combinations. The latter was chosen.

Port_Quarter_Strata_temp<-paste(df1$onShoreSampLoc,df1$onShoreSampQuarter)
weight.strata<-sort(tapply(df1$landWt, Port_Quarter_Strata_temp, sum), decreasing=TRUE)

# raw plots
windows(15,7)
par(mfrow=c(1,2))
```

```
  barplot(weight.strata[1:40]/1000, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata), xlab = "Ports")

  dev.off()
  round(cumsum(weight.strata)/sum(weight.strata)*100,0)

  # seleccao de main_strata_regional portos*Quarter + OTHER # 36%
   # provavelmente pouco representativa
   if(allocType == "perc_land")
    {
    n_main_strata_regional<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional))+1
    }
   percland.main.strata<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_regional]
   percland.main.strata

   df1$PortQuarter2c<-NA
   df1$PortQuarter2c[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_regional]]<-
Port_Quarter_Strata_temp[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_regional]]
   df1$PortQuarter2c[!Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_regional]]<-"OT999"

  # barplots (landings)
   par(mfrow=c(2,2))
   weight.strata<-sort(tapply(df1$landWt, df1$PortQuarter2c, sum)/1000, decreasing=TRUE)
   cores<-str_sub(names(weight.strata),1,2)
   cores[cores=="PT"]<-"red"
   cores[cores=="ES"]<-"blue"
   cores[cores=="BC"]<-"green"
   cores[cores=="OT"]<-"gray"
   barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports*Quarters (colour = Institute)")
   a<-tapply(df1$landWt, str_sub(df1$PortQuarter2c,1,2), sum)
   barplot(a/1000, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = "Landings
(ton)", xlab = "Institute")

   # distribution by quarters
    weight.strata2<-sort(tapply(df1$landWt, df1$PortQuarter2c, sum)/1000, decreasing=TRUE)
    names(weight.strata2)[grepl(names(weight.strata), pat="Q1")]<-"Q1"
    names(weight.strata2)[grepl(names(weight.strata), pat="Q2")]<-"Q2"
    names(weight.strata2)[grepl(names(weight.strata), pat="Q3")]<-"Q3"
    names(weight.strata2)[grepl(names(weight.strata), pat="Q4")]<-"Q4"

    weight.strata2<-tapply(weight.strata2, names(weight.strata2), sum)

    cores<-names(weight.strata2)
    cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
    cores[grepl(cores, pat="Q1")]<-"red"
    cores[grepl(cores, pat="Q2")]<-"blue"
    cores[grepl(cores, pat="Q3")]<-"green"
    cores[grepl(cores, pat="Q4")]<-"pink"
    barplot(weight.strata2, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Quarters")

   # distribution by ports
    cores<-names(weight.strata)
    cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
    cores[grepl(cores, pat="Q1")]<-"red"
    cores[grepl(cores, pat="Q2")]<-"blue"
    cores[grepl(cores, pat="Q3")]<-"green"
    cores[grepl(cores, pat="Q4")]<-"pink"
    barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports*Quarters (colour = Quarter)")

    title(main=c("SHake (All Vessels Setting) - Scenario 2c"), outer=TRUE, line=-1)
    title(main=paste("main strata (",length(weight.strata[!grepl(names(weight.strata), pat="999")]),") =",percland.main.strata,"% kg; all
strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)

   savePlot(file="Results\\Shake_AllVessels_Sc2c_landings.png",type="png")

  # sample allocation
   samples<-round(prop.table(tapply(df1$landWt, df1$PortQuarter2c, sum))*psuALL,0) # 2930

  # barplots (psu)
   windows(15,7)
   par(mfrow=c(2,2))
```

```
   cores<-str_sub(names(samples),1,2)
   cores[cores=="PT"]<-"red"
   cores[cores=="ES"]<-"blue"
   cores[cores=="BC"]<-"green"
   cores[grepl(names(samples), pat="999")]<-"gray"
   barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports*Quarters (colour = Institute)")
   a<-tapply(samples, str_sub(names(samples),1,2), sum)
   barplot(a, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = psuType, xlab =
"Institute")

   # distribution by quarters
    samples2<-samples
    names(samples2)[grepl(names(samples2), pat="Q1")]<-"Q1"
    names(samples2)[grepl(names(samples2), pat="Q2")]<-"Q2"
    names(samples2)[grepl(names(samples2), pat="Q3")]<-"Q3"
    names(samples2)[grepl(names(samples2), pat="Q4")]<-"Q4"

    samples2<-tapply(samples2, names(samples2), sum)

    cores<-names(samples2)
    cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
    cores[grepl(cores, pat="Q1")]<-"red"
    cores[grepl(cores, pat="Q2")]<-"blue"
    cores[grepl(cores, pat="Q3")]<-"green"
    cores[grepl(cores, pat="Q4")]<-"pink"
    barplot(samples2, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Quarters")

   # distribution by Quarters and ports
    cores<-names(samples)
    cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
    cores[grepl(cores, pat="Q1")]<-"red"
    cores[grepl(cores, pat="Q2")]<-"blue"
    cores[grepl(cores, pat="Q3")]<-"green"
    cores[grepl(cores, pat="Q4")]<-"pink"
    barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports*Quarters (colour = Quarter)")

    title(main="Sample Distribution SHake (All Vessels Setting) - Scenario 2c", outer=TRUE, line=-1)
    title(main=paste("(strata <2 samples = ", sum(samples<2),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
pat="999")]),") =",percland.main.strata,"% kg; all strata = 100% kg)"), outer=TRUE, line=-2, cex.main=0.9)
    savePlot(file="Results\\Shake_AllVessels_Sc2c_samples.png",type="png")

   # csv ouput
    res<-melt(samples)
    colnames(res)[1]<-"PortQuarter2c"
    colnames(res)[2]<-"n_psu"
    res$n_ssu_per_psu=n_ssu_per_psu
    res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata; sum(res[,2]); sum(res[,2]*res[,3])
    write.csv2(res,"Results\\Shake_AllVessels_Sc2c_samples.csv", row.names=FALSE)

    graphics.off()

# =====================================
# Scenario 3 [countries as main strata]
# =====================================

 # =====================================
 # Country/Institute strata, sample allocation fixed(Ctr30)
 # =====================================

   res<-data.frame(COUNTRY = c("ES", "PT", "BC"), n_psu=c(psuES,psuPT,psuBC), n_ssu_per_psu=n_ssu_per_psu)
   write.csv2(res,"Results\\Shake_AllVessels_Sc30p_samples.csv", row.names=FALSE)


 # =====================================
 # Country/Institute strata, sample allocation proportional to landings (Ctr30)
 # =====================================

  weight.strata<-sort(tapply(df1$landWt, df1$onShoreSampCtr, sum), decreasing=TRUE)

  # raw plots
   windows(15,7)
```

```
  par(mfrow=c(1,2))
  barplot(weight.strata/1000, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata/1000))
  dev.off()

  weight.strata[1:25]/1000
  round(cumsum(weight.strata)/sum(weight.strata)*100,0) [1:25]

 #seleccao de n_main_strata_regional portos + OTHER #
  # note: BC disappears
  if(allocType == "perc_land")
   {
   n_main_strata_regional<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional))+1
   }
  percland.main.strata<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_regional]
  percland.main.strata

  df1$Ctr30<-NA
  df1$Ctr30[df1$onShoreSampCtr %in% names(weight.strata)[1:n_main_strata_regional]]<-df1$onShoreSampCtr[df1$onShoreSampCtr
%in% names(weight.strata)[1:n_main_strata_regional]]
  df1$Ctr30[!df1$onShoreSampCtr %in% names(weight.strata)[1:n_main_strata_regional]]<-"OT999"

 # barplot (landings)
 windows(7.5,7)
  par(mfrow=c(1,1))
  weight.strata<-sort(tapply(df1$landWt, df1$Ctr30, sum)/1000, decreasing=TRUE)
  cores<-str_sub(names(weight.strata),1,2)
  cores[cores=="PT"]<-"red"
  cores[cores=="ES"]<-"blue"
  cores[cores=="BC"]<-"green"
  cores[cores=="OT"]<-"gray"
  barplot(weight.strata, col=cores, names.arg = c("IEO","IPMA","MIX"),las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Institute")
  a<-tapply(df1$landWt, str_sub(df1$Ctr30,1,2), sum)
  #barplot(a/1000, col=c("blue","gray","red"), las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Country")
  title(main="SHake (All Vessels Setting) - Scenario 30", outer=TRUE, line=-1)
  title(main=paste("main strata (",length(weight.strata[!grepl(names(weight.strata), pat="999")]),") =",percland.main.strata,"% kg; all
strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
  savePlot(file="Results\\Shake_AllVessels_Sc30_landings.png",type="png")

 # sample allocation
  samples<-round(prop.table(tapply(df1$landWt, df1$Ctr30, sum))*psuALL,0) # 3792

 # final barplots (trips to sample)
  windows(15,7)
  par(mfrow=c(1,1))
  cores<-str_sub(names(samples),1,2)
  cores[cores=="PT"]<-"red"
  cores[cores=="ES"]<-"blue"
  cores[cores=="BC"]<-"green"
  cores[grepl(names(samples), pat="999")]<-"gray"
  barplot(samples, col=cores, names.arg = c("IEO","MIX","IPMA"),las=2, cex.names=0.5, ylab = psuType, xlab = "Country")
  #a<-tapply(samples, str_sub(names(samples),1,2), sum)
 # barplot(a, col=c("blue","gray","red"), las=2, cex.names=0.5, ylab = psuType, xlab = "Institute")
  title(main="Sample Distribution SHake (All Vessels Setting)  - Scenario 30", outer=TRUE, line=-1)
  title(main=paste("(strata <2 samples = ", sum(samples<1),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
pat="999")]),") =",percland.main.strata,"% kg; all strata = 100% kg)"), outer=TRUE, line=-2, cex.main=0.9)
  savePlot(file="Results\\Shake_AllVessels_Sc30_samples.png",type="png")

 # csv ouput
  res<-melt(samples)
  colnames(res)[1]<-"Ctr30"
  colnames(res)[2]<-"n_psu"
  res$n_ssu_per_psu=n_ssu_per_psu
  res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata
  write.csv2(res,"Results\\Shake_AllVessels_Sc30_samples.csv", row.names=FALSE)

 graphics.off()


# ====================================
# Port strata 2013 within Country/Institute (3a)
```

```
# ====================================

 # PRT strata
  df1pt<-df1[df1$onShoreSampCtr=="PT",]

  weight.strata<-sort(tapply(df1pt$landWt, df1pt$onShoreSampLoc, sum), decreasing=TRUE)

 # raw plots
  windows(15,7)
  par(mfrow=c(1,2))
  barplot(weight.strata, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata))
  dev.off()
  round(cumsum(weight.strata)/sum(weight.strata)*100,0)

  # seleccao de main_strata_national_pt portos + OTHER # 86%
  if(allocType == "perc_land")
  {
  n_main_strata_national_pt<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional_pt))+1
  }

  percland.main.strata_pt<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_pt]
  percland.main.strata_pt

  df1pt$Port3a<-NA
  df1pt$Port3a[df1pt$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_pt]]<-
df1pt$onShoreSampLoc[df1pt$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_pt]]
  df1pt$Port3a[!df1pt$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_pt]]<-"PT999"

 # ESP strata
  df1es<-df1[df1$onShoreSampCtr=="ES",]

  weight.strata<-sort(tapply(df1es$landWt, df1es$onShoreSampLoc, sum), decreasing=TRUE)

  # raw plots
  windows(15,7)
  par(mfrow=c(1,2))
  barplot(weight.strata[1:40]/1000, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata))
  dev.off()
  if(allocType == "perc_land")
  {
  n_main_strata_national_es<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional_es))+1
  }
  percland.main.strata_es<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_es]
  percland.main.strata_es

  # seleccao de 9 portos + OTHER # 58%
  round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_es]

  df1es$Port3a<-NA
  df1es$Port3a[df1es$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_es]]<-
df1es$onShoreSampLoc[df1es$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_es]]
  df1es$Port3a[!df1es$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_es]]<-"ES999"

 # BC strata
  df1bc<-df1[df1$onShoreSampCtr=="BC",]

  weight.strata<-sort(tapply(df1bc$landWt, df1bc$onShoreSampLoc, sum), decreasing=TRUE)

  # raw plots
  windows(15,7)
  par(mfrow=c(1,2))
  barplot(weight.strata[1:40]/1000, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata))
  dev.off()
  if(allocType == "perc_land")
  {
    n_main_strata_national_bc<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional_bc))+1
  }
  percland.main.strata_bc<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_bc]
```

```
    percland.main.strata_bc

    # seleccao de 9 portos + OTHER # 58%
    round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_bc]

    df1bc$Port3a<-NA
    df1bc$Port3a[df1bc$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_bc]]<-
  df1bc$onShoreSampLoc[df1bc$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_bc]]
    df1bc$Port3a[!df1bc$onShoreSampLoc %in% names(weight.strata)[1:n_main_strata_national_bc]]<-"BC999"

   # OUT_OF_FRAME strata
    df1out<-df1[! df1$onShoreSampCtr %in% c("PT","ES","BC"),]
    df1out$Port3a<-paste("OT",999,sep="")

   # reconstructs original data (joins PT+ES+BC+OUT)
    df1<-rbind(df1pt, df1es, df1bc, df1out)

   # final barplot
    windows(15,7)
    par(mfrow=c(1,2))
    weight.strata<-sort(tapply(df1$landWt, df1$Port3a, sum)/1000, decreasing=TRUE)
    cores<-str_sub(names(weight.strata),1,2)
    cores[cores=="PT"]<-"red"
    cores[cores=="ES"]<-"blue"
    cores[cores=="BC"]<-"green"
    cores[grepl(names(weight.strata), pat="999")]<-"gray"
    barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
    a<-tapply(df1$landWt, str_sub(df1$Port3a,1,2), sum)
    barplot(a/1000, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = "Landings
  (ton)", xlab = "Institute")
    title(main=c("SHake (All Vessels Setting) - Scenario 3a"), outer=TRUE, line=-1)
    percland.main.strata_all<-round(sum(weight.strata[!grepl(names(weight.strata), pat="999")])/sum(weight.strata)*100,0)
    title(main=paste("main strata (",length(weight.strata[!grepl(names(weight.strata),
  pat="999")]),"=",n_main_strata_national_pt,"+",n_main_strata_national_es,"+",n_main_strata_national_bc,") =",
         percland.main.strata_pt,"% kg (pt);", percland.main.strata_es,"% kg (es);", percland.main.strata_bc,"% kg (bc);",
         percland.main.strata_all,"% kg (pt+es+bc); all strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
    savePlot(file="Results\\Shake_AllVessels_Sc3a_landings.png",type="png")

   # sample allocation
    samples<-tapply(df1$landWt, list(df1$Port3a, str_sub(df1$Port3a,1,2)), sum, na.rm=T)
    samples[is.na(samples)]<-0
    samples<-prop.table(samples,2)
    samples[,"PT"]<-round(samples[,"PT"]*psuPT)
    samples[,"ES"]<-round(samples[,"ES"]*psuES)
    samples[,"BC"]<-round(samples[,"BC"]*psuBC)
    samples[,"OT"]<-round(samples[,"OT"]*psuOT)
    samples<-margin.table(samples,1); sum(samples)

    # final barplots
    windows(15,7)
    par(mfrow=c(1,2))
    cores<-str_sub(names(samples),1,2)
    cores[cores=="PT"]<-"red"
    cores[cores=="ES"]<-"blue"
    cores[cores=="BC"]<-"green"
    cores[grepl(names(samples), pat="999")]<-"gray"
    barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports")
    a<-tapply(samples, str_sub(names(samples),1,2), sum)
    barplot(a, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = psuType, xlab =
  "Institute")
    title(main="SHake (All Vessels Fleet Setting) Hake Fleet Setting - Scenario 3a", outer=TRUE, line=-1)

    title(main=paste("strata <2 samples = ", sum(samples<2),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
  pat="999")]),"=",
    n_main_strata_national_pt,"+",n_main_strata_national_es,"+",n_main_strata_national_bc,") =",
    percland.main.strata_pt,"% kg (pt);", percland.main.strata_es,"% kg (es);",percland.main.strata_bc,"% kg (bc);",
    percland.main.strata_all,"% kg (pt+es+bc); all strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
    savePlot(file="Results\\Shake_AllVessels_Sc3a_samples.png",type="png")

    # csv ouput
    res<-melt(samples)
```

```
 colnames(res)[1]<-"Port3a"
 colnames(res)[2]<-"n_psu"
 res$n_ssu_per_psu=n_ssu_per_psu
 res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata
 write.csv2(res,"Results\\Shake_AllVessels_Sc3a_samples.csv", row.names=FALSE)

graphics.off()

# ====================================
# Quarter strata 2013 within Country/Institute (3b)
# ====================================

# PRT strata
 df1pt<-df1[df1$onShoreSampCtr=="PT",]

 weight.strata<-sort(tapply(df1pt$landWt, df1pt$onShoreSampQuarter, sum), decreasing=TRUE)

# raw plots
 windows(15,7)
 par(mfrow=c(1,2))
 barplot(weight.strata, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
 plot(ecdf(weight.strata))
 dev.off()
 round(cumsum(weight.strata)/sum(weight.strata)*100,0)
 percland.main.strata_pt<-100

# choice: all quarters are to be sampled independently
 df1pt$Quarter3b<-paste("PT",df1pt$onShoreSampQuarter,sep="")

# ESP strata
 df1es<-df1[df1$onShoreSampCtr=="ES",]

 weight.strata<-sort(tapply(df1es$landWt, df1es$onShoreSampQuarter, sum), decreasing=TRUE)

 # raw plots
 windows(15,7)
 par(mfrow=c(1,2))
 barplot(weight.strata, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
 plot(ecdf(weight.strata))
 dev.off()
 round(cumsum(weight.strata)/sum(weight.strata)*100,0)

 percland.main.strata_es<-100
 percland.main.strata_es

 # choice: all quarters are to be sampled independently
 df1es$Quarter3b<-paste("ES",quarters(df1es$arvDate),sep="")

# BC strata
 df1bc<-df1[df1$onShoreSampCtr=="BC",]

 weight.strata<-sort(tapply(df1bc$landWt, df1bc$onShoreSampQuarter, sum), decreasing=TRUE)

# raw plots
 windows(15,7)
 par(mfrow=c(1,2))
 barplot(weight.strata, las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
 plot(ecdf(weight.strata))
 dev.off()
 round(cumsum(weight.strata)/sum(weight.strata)*100,0)

 percland.main.strata_bc<-100
 percland.main.strata_bc

# choice: all quarters are to be sampled independently
 df1bc$Quarter3b<-paste("BC",quarters(df1bc$arvDate),sep="")

# OUT_OF_FRAME strata
 df1out<-df1[! df1$onShoreSampCtr %in% c("PT","ES","BC"),]
 df1out$Quarter3b<-paste("OT",999,sep="")
```

```
  # reconstructs original data
   df1<-rbind(df1pt, df1es, df1bc,df1out)

  # final barplot
   windows(15,7)
   par(mfrow=c(1,2))
   weight.strata<-sort(tapply(df1$landWt, df1$Quarter3b, sum)/1000, decreasing=TRUE)
   cores<-str_sub(names(weight.strata),1,2)
   cores[cores=="PT"]<-"red"
   cores[cores=="ES"]<-"blue"
   cores[cores=="BC"]<-"green"
   cores[grepl(names(weight.strata), pat="999")]<-"gray"
   barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
   a<-tapply(df1$landWt, str_sub(df1$Quarter3b,1,2), sum)
   barplot(a/1000, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = "Landings
(ton)", xlab = "Institute")
   title(main=c("SHake (All Vessels Setting) - Scenario 3b"), outer=TRUE, line=-1)
   percland.main.strata_all<-round(sum(weight.strata[!grepl(names(weight.strata), pat="999")])/sum(weight.strata)*100,0)
   title(main=paste("main strata (",length(weight.strata[!grepl(names(weight.strata), pat="999")]),") =",
       percland.main.strata_pt,"% kg (pt);", percland.main.strata_es,"% kg (es);", percland.main.strata_bc,"% kg (bc);",
       percland.main.strata_all,"% kg (pt+es+bc); all strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
   savePlot(file="Results\\Shake_AllVessels_Sc3b_landings.png",type="png")

    # sample allocation
    samples<-tapply(df1$landWt, list(df1$Quarter3b, str_sub(df1$Quarter3b,1,2)), sum, na.rm=T)
    samples[is.na(samples)]<-0
    samples<-prop.table(samples,2)
    samples[,"PT"]<-round(samples[,"PT"]*psuPT)
    samples[,"ES"]<-round(samples[,"ES"]*psuES)
    samples[,"BC"]<-round(samples[,"BC"]*psuBC)
    samples[,"OT"]<-round(samples[,"OT"]*psuOT)
    samples<-margin.table(samples,1); sum(samples)

   # final barplots
    windows(15,7)
    par(mfrow=c(1,2))
    cores<-str_sub(names(samples),1,2)
    cores[cores=="PT"]<-"red"
    cores[cores=="ES"]<-"blue"
    cores[cores=="BC"]<-"green"
    cores[grepl(names(samples), pat="999")]<-"gray"
    barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports")
    a<-tapply(samples, str_sub(names(samples),1,2), sum)
    barplot(a, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = psuType, xlab =
"Institute")
    title(main="SHake (All Vessels Setting) Hake Fleet Setting - Scenario 3b", outer=TRUE, line=-1)

    title(main=paste("strata <2 samples = ", sum(samples<2),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
pat="999")]),") =",
       percland.main.strata_pt,"% kg (pt);", percland.main.strata_es,"% kg (es);",percland.main.strata_bc,"% kg (bc);",
       percland.main.strata_all,"% kg (pt+es+bc); all strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
    savePlot(file="Results\\Shake_AllVessels_Sc3b_samples.png",type="png")

    # csv ouput
    res<-melt(samples)
    colnames(res)[1]<-"Quarter3b"
    colnames(res)[2]<-"n_psu"
    res$n_ssu_per_psu=n_ssu_per_psu
    res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata
    write.csv2(res,"Results\\Shake_AllVessels_Sc3b_samples.csv", row.names=FALSE)

   graphics.off()

 # ====================================
 # Port*Quarter strata 2013 within Country/Institute (3c)
 # ====================================

  # Note: Two ways were considered: a) use the 'major' ports previously selected and b) use all port*quarter combinations. The latter was
chosen.

  # PRT strata
```

```
  df1pt<-df1[df1$onShoreSampCtr=="PT",]

  Port_Quarter_Strata_temp<-paste(df1pt$onShoreSampLoc,df1pt$onShoreSampQuarter)

  windows(15,7)
  par(mfrow=c(1,2))
  weight.strata<-sort(tapply(df1pt$landWt, Port_Quarter_Strata_temp, sum), decreasing=TRUE)
  barplot(weight.strata[1:40], las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata))
  dev.off()

  round(cumsum(weight.strata)/sum(weight.strata)*100,0)

  if(allocType == "perc_land")
   {
   n_main_strata_national_pt<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional_pt))+1
   }
   percland.main.strata_pt<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_pt]
   percland.main.strata_pt

  # seleccao de 9 portos*Quarter + OTHER
   # provavelmente pouco representativa
  round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_pt]

  df1pt$PortQuarter3c<-NA
  df1pt$PortQuarter3c[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_pt]]<-
Port_Quarter_Strata_temp[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_pt]]
  df1pt$PortQuarter3c[!Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_pt]]<-"PT999"

 # ESP strata

  df1es<-df1[df1$onShoreSampCtr=="ES",]
  Port_Quarter_Strata_temp<-paste(df1es$onShoreSampLoc,df1es$onShoreSampQuarter)
  weight.strata<-sort(tapply(df1es$landWt, Port_Quarter_Strata_temp, sum), decreasing=TRUE)

  windows(15,7)
  par(mfrow=c(1,2))
  barplot(weight.strata[1:40], las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
  plot(ecdf(weight.strata))
  dev.off()

  round(cumsum(weight.strata)/sum(weight.strata)*100,0)

  if(allocType == "perc_land")
   {
   n_main_strata_national_es<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional_es))+1
   }
  percland.main.strata_es<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_es]
  percland.main.strata_es

  # seleccao de n_main_strata_national_es portos*Quarter + OTHER
  # provavelmente pouco representativa
  round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_es]

  df1es$PortQuarter3c<-NA
  df1es$PortQuarter3c[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_es]]<-
Port_Quarter_Strata_temp[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_es]]
  df1es$PortQuarter3c[!Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_es]]<-"ES999"


 # BC strata

 df1bc<-df1[df1$onShoreSampCtr=="BC",]
 Port_Quarter_Strata_temp<-paste(df1bc$onShoreSampLoc,df1bc$onShoreSampQuarter)

 weight.strata<-sort(tapply(df1bc$landWt, Port_Quarter_Strata_temp, sum), decreasing=TRUE)

 windows(15,7)
 par(mfrow=c(1,2))
 barplot(weight.strata[1:40], las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports")
```

```
  plot(ecdf(weight.strata))
  dev.off()

  round(cumsum(weight.strata)/sum(weight.strata)*100,0)

  if(allocType == "perc_land")
  {
  n_main_strata_national_bc<-length(which(round(cumsum(weight.strata)/sum(weight.strata)*100,0)<perc_land_regional_bc))+1
  }
  percland.main.strata_bc<-round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_bc]
  percland.main.strata_bc

  # seleccao de n_main_strata_national_bc portos*Quarter + OTHER
  # provavelmente pouco representativa
  round(cumsum(weight.strata)/sum(weight.strata)*100,0)[n_main_strata_national_bc]

  df1bc$PortQuarter3c<-NA
  df1bc$PortQuarter3c[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_bc]]<-
Port_Quarter_Strata_temp[Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_bc]]
  df1bc$PortQuarter3c[!Port_Quarter_Strata_temp %in% names(weight.strata)[1:n_main_strata_national_bc]]<-"BC999"

  # OUT_OF_FRAME strata
   df1out<-df1[! df1$onShoreSampCtr %in% c("PT","ES","BC"),]
   df1out$PortQuarter3c<-"OT999"

  # reconstructs original data
   df1<-rbind(df1pt, df1es, df1bc, df1out)

  # final barplot
  windows(15,7)
  par(mfrow=c(2,2))
  weight.strata<-sort(tapply(df1$landWt, df1$PortQuarter3c, sum)/1000, decreasing=TRUE)
  cores<-str_sub(names(weight.strata),1,2)
  cores[cores=="PT"]<-"red"
  cores[cores=="ES"]<-"blue"
  cores[cores=="BC"]<-"green"
  cores[grepl(names(weight.strata), pat="999")]<-"gray"
  barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports*Quarters (colour: Institute)")
  a<-tapply(df1$landWt, str_sub(df1$PortQuarter3c,1,2), sum)
  barplot(a/1000, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = "Landings
(ton)", xlab = "Institute")
  title(main=c("SHake (All Vessels Setting) Hake Fleet Setting - Scenario 3c"), outer=TRUE, line=-1)

  # distribution by quarters
  weight.strata2<-sort(tapply(df1$landWt, df1$PortQuarter3c, sum)/1000, decreasing=TRUE)
  names(weight.strata2)[grepl(names(weight.strata), pat="Q1")]<-"Q1"
  names(weight.strata2)[grepl(names(weight.strata), pat="Q2")]<-"Q2"
  names(weight.strata2)[grepl(names(weight.strata), pat="Q3")]<-"Q3"
  names(weight.strata2)[grepl(names(weight.strata), pat="Q4")]<-"Q4"

   weight.strata2<-tapply(weight.strata2, names(weight.strata2), sum)

   cores<-names(weight.strata2)
   cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
   cores[grepl(cores, pat="Q1")]<-"red"
   cores[grepl(cores, pat="Q2")]<-"blue"
   cores[grepl(cores, pat="Q3")]<-"green"
   cores[grepl(cores, pat="Q4")]<-"pink"
   barplot(weight.strata2, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Quarters")

  # distribution by ports*quarters
  cores<-names(weight.strata)
  cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
  cores[grepl(cores, pat="Q1")]<-"red"
  cores[grepl(cores, pat="Q2")]<-"blue"
  cores[grepl(cores, pat="Q3")]<-"green"
  cores[grepl(cores, pat="Q4")]<-"pink"
  barplot(weight.strata, col=cores,las=2, cex.names=0.5, ylab = "Landings (ton)", xlab = "Ports*Quarters  (colour: Quarters)")

  percland.main.strata_all<-round(sum(weight.strata[!grepl(names(weight.strata), pat="999")])/sum(weight.strata)*100,0)
  title(main=paste("main strata (",length(weight.strata[!grepl(names(weight.strata), pat="999")]),"=",
```

```
   n_main_strata_national_pt,"+",n_main_strata_national_es,"+",n_main_strata_national_bc,") =",
   percland.main.strata_pt,"% kg (pt);", percland.main.strata_es,"% kg (es);",percland.main.strata_bc,"% kg (bc);",
   percland.main.strata_all,"% kg (pt+es+bc); all strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
  savePlot(file="Results\\Shake_AllVessels_Sc3c_landings.png",type="png")

  # sample allocation
  samples<-tapply(df1$landWt, list(df1$PortQuarter3c, str_sub(df1$PortQuarter3c,1,2)), sum, na.rm=T)
  samples[is.na(samples)]<-0
  samples<-prop.table(samples,2)
  samples[,"PT"]<-round(samples[,"PT"]*psuPT)
  samples[,"ES"]<-round(samples[,"ES"]*psuES)
  samples[,"BC"]<-round(samples[,"BC"]*psuBC)
  samples[,"OT"]<-round(samples[,"OT"]*psuOT)
  samples<-margin.table(samples,1); sum(samples)

  # final barplots
   windows(15,7)
   par(mfrow=c(2,2))
   cores<-str_sub(names(samples),1,2)
   cores[cores=="PT"]<-"red"
   cores[cores=="ES"]<-"blue"
   cores[cores=="BC"]<-"green"
   cores[grepl(names(samples), pat="999")]<-"gray"
   barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports*Quarters (colour: Institute)")
   a<-tapply(samples, str_sub(names(samples),1,2), sum)
   barplot(a, col=c("green", "blue","gray","red"), names.arg = c("AZTI","IEO","MIX","IPMA"), las=2, cex.names=0.5, ylab = psuType, xlab =
"Institute")

   # distribution by quarters
   samples2<-samples
   names(samples2)[grepl(names(samples2), pat="Q1")]<-"Q1"
   names(samples2)[grepl(names(samples2), pat="Q2")]<-"Q2"
   names(samples2)[grepl(names(samples2), pat="Q3")]<-"Q3"
   names(samples2)[grepl(names(samples2), pat="Q4")]<-"Q4"

    samples2<-tapply(samples2, names(samples2), sum)

    cores<-names(samples2)
    cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
    cores[grepl(cores, pat="Q1")]<-"red"
    cores[grepl(cores, pat="Q2")]<-"blue"
    cores[grepl(cores, pat="Q3")]<-"green"
    cores[grepl(cores, pat="Q4")]<-"pink"
    barplot(samples2, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Quarters")

   # distribution by Quarters and ports
    cores<-names(samples)
    cores[!grepl(cores, pat="Q1") & !grepl(cores, pat="Q2") & !grepl(cores, pat="Q3") & !grepl(cores, pat="Q4")]<-"gray"
    cores[grepl(cores, pat="Q1")]<-"red"
    cores[grepl(cores, pat="Q2")]<-"blue"
    cores[grepl(cores, pat="Q3")]<-"green"
    cores[grepl(cores, pat="Q4")]<-"pink"
    barplot(samples, col=cores,las=2, cex.names=0.5, ylab = psuType, xlab = "Ports*Quarters (colour: Quarters)")

   title(main="SHake (All Vessels Setting) - Scenario 3c", outer=TRUE, line=-1)
   title(main=paste("strata <2 samples = ", sum(samples<2),"; main strata (",length(weight.strata[!grepl(names(weight.strata),
pat="999")]),"=",
     n_main_strata_national_pt,"+",n_main_strata_national_es,"+",n_main_strata_national_bc,") =",
     percland.main.strata_pt,"% kg (pt);", percland.main.strata_es,"% kg (es);",percland.main.strata_bc,"% kg (bc);",
     percland.main.strata_all,"% kg (pt+es+bc); all strata = 100% kg"), outer=TRUE, line=-2, cex.main=0.9)
   savePlot(file="Results\\Shake_AllVessels_Sc3c_samples.png",type="png")

   # csv ouput
   res<-melt(samples)
   colnames(res)[1]<-"PortQuarter3c"
   colnames(res)[2]<-"n_psu"
   res$n_ssu_per_psu=n_ssu_per_psu
   res[res[,2]<min_psu_per_strata,2]<-min_psu_per_strata
   write.csv2(res,"Results\\Shake_AllVessels_Sc3c_samples.csv", row.names=FALSE)

   graphics.off()
```

```
# ====================================
# Scenario 4
# ====================================

 # Not developed


# ====================================
# Updates db with stratification fields and handles NA
# ====================================

 # subsets simulation dataset from original data
  df2<-droplevels(db1[format(db1$arvDate, "%Y" )=="2014",])

 # adds "Port2a"

  # determines unique combinations in planning dataset
  temp<-unique(df1[c("onShoreSampLoc","Port2a")])

  # merges them into simulation dataset
  df2<-merge(df2, temp, by="onShoreSampLoc", all.x=T)

  # displays final situation (note: NAs are outofframe - they were not present in planning data)
   head(table(df2$onShoreSampLoc, df2$Port2a, useNA="al"))

 # adds "Quarter2b"

  # determines unique combinations in planning dataset
  temp<-unique(df1[c("onShoreSampQuarter","Quarter2b")])

  # merges them into simulation dataset
  df2<-merge(df2, temp, by="onShoreSampQuarter", all.x=T)

  # displays final situation (note: NAs are outofframe - they were not present in planning data)
  table(df2$onShoreSampQuarter, df2$Quarter2b, useNA="al")

 # adds "PortQuarter2c"

  # determines unique combinations in planning dataset and creates ID
  temp<-unique(df1[c("onShoreSampLoc","onShoreSampQuarter","PortQuarter2c")])
  temp$ID<-paste(temp$onShoreSampLoc,temp$onShoreSampQuarter)

  # creates id in simulation dataset
  df2$ID<-paste(df2$onShoreSampLoc,df2$onShoreSampQuarter)

  # merges strata into the simulation dataset
  df2<-merge(df2, temp[c("ID","PortQuarter2c")], by="ID", all.x=T)

  # displays final situation (note: NAs are outofframe - they were not present in planning data)
  head(table(df2$ID, df2$PortQuarter2c, useNA="al"))
  margin.table(table(df2$ID, df2$PortQuarter2c, useNA="al"),2)
  unique(df2$ID[is.na(df2$PortQuarter2c)])

 # adds "Ctr30"

  # determines unique combinations in planning dataset
  temp<-unique(df1[c("onShoreSampCtr","Ctr30")])

  # merges them into simulation dataset
  df2<-merge(df2, temp, by="onShoreSampCtr", all.x=T)

  # displays final situation (note: NAs are outofframe - they were not present in planning data)
  head(table(df2$onShoreSampCtr, df2$Ctr30, useNA="al"))

 # adds "Port3a"

  # determines unique combinations in planning dataset
  temp<-unique(df1[c("onShoreSampCtr","onShoreSampLoc","Port3a")])
  temp$ID<-paste(temp$onShoreSampCtr,temp$onShoreSampLoc)
```

```
 # creates id in simulation dataset
 df2$ID<-paste(df2$onShoreSampCtr,df2$onShoreSampLoc)

 # merges strata into the simulation dataset
 df2<-merge(df2, temp[c("ID","Port3a")], by="ID", all.x=T)

 # displays final situation (note: NAs are outofframe - they were not present in planning data)
 head(table(df2$ID, df2$Port3a, useNA="al"))
 margin.table(table(df2$ID, df2$Port3a, useNA="al"),2)
 unique(df2$ID[is.na(df2$Port3a)])

# adds "Quarter3b"

 # determines unique combinations in planning dataset
 temp<-unique(df1[c("onShoreSampCtr","onShoreSampQuarter","Quarter3b")])
 temp$ID<-paste(temp$onShoreSampCtr,temp$onShoreSampQuarter)

 # creates id in simulation dataset
 df2$ID<-paste(df2$onShoreSampCtr,df2$onShoreSampQuarter)

 # merges strata into the simulation dataset
 df2<-merge(df2, temp[c("ID","Quarter3b")], by="ID", all.x=T)

 # displays final situation and corrects NAs (note: NAs are outofframe - they were not present in planning data)
 head(table(df2$ID, df2$Quarter3b, useNA="al"))
 margin.table(table(df2$ID, df2$Quarter3b, useNA="al"),2)
 unique(df2$ID[is.na(df2$Quarter3b)])

# adds "PortQuarter3c"

 # determines unique combinations in planning dataset
 temp<-unique(df1[c("onShoreSampCtr","onShoreSampLoc","onShoreSampQuarter","PortQuarter3c")])
 temp$ID<-paste(temp$onShoreSampCtr,temp$onShoreSampLoc,temp$onShoreSampQuarter)

 # creates id in simulation dataset
 df2$ID<-paste(df2$onShoreSampCtr,df2$onShoreSampLoc,df2$onShoreSampQuarter)

 # merges strata into the simulation dataset
 df2<-merge(df2, temp[c("ID","PortQuarter3c")], by="ID", all.x=T)

 # displays final situation and corrects NAs (note: NAs are outofframe - they were not present in planning data)
 head(table(df2$ID, df2$PortQuarter3c, useNA="al"))
 margin.table(table(df2$ID, df2$PortQuarter3c, useNA="al"),2)
 unique(df2$ID[is.na(df2$PortQuarter3c)])

# handling of NAs
 # NAs in 2014 correspond to Ports, Quarters, or PortQuarters where there was no landings (of anything!) in 2013
 # because we are emulating the sampling design for 2014 at end of 2013, these Ports, Quarters, or Port*Quarters are out of our sampling
frame for 2014
  # to avoid this, we could consider a frame with all potential ports and all quarters, irrespectively of whether or not they registered any
  # landings (hake or other species). This will however be inneficient and no one really would do that unless landing sites were really
unpredictable.
  # impact analyses confirm that they are predictable and that OUT_OF_FRAME situations are minimal (see below)

 df2$Port2a[is.na(df2$Port2a)]<-"OUT_OF_FRAME"
 df2$Quarter2b[is.na(df2$Quarter2b)]<-"OUT_OF_FRAME"
 df2$PortQuarter2c[is.na(df2$PortQuarter2c)]<-"OUT_OF_FRAME"
 df2$Ctr30[is.na(df2$Ctr30)]<-paste(df2$onShoreSampCtr[is.na(df2$Port3a)],"OUT_OF_FRAME")
 df2$Port3a[is.na(df2$Port3a)]<-paste(df2$onShoreSampCtr[is.na(df2$Port3a)],"OUT_OF_FRAME")
 df2$Quarter3b[is.na(df2$Quarter3b)]<-paste(df2$onShoreSampCtr[is.na(df2$Quarter3b)],"OUT_OF_FRAME")
 df2$PortQuarter3c[is.na(df2$PortQuarter3c)]<-paste(df2$onShoreSampCtr[is.na(df2$PortQuarter3c)],"OUT_OF_FRAME")

 # for display only: % of 2014 landings OUT_OF_FRAME under the different scenarios (very low)
 round(sum(df2[grep(df2$Port2a, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
 round(sum(df2[grep(df2$Quarter2b, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
 round(sum(df2[grep(df2$PortQuarter2c, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
 round(sum(df2[grep(df2$Ctr30, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
 round(sum(df2[grep(df2$Port3a, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
 round(sum(df2[grep(df2$Quarter3b, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
 round(sum(df2[grep(df2$PortQuarter3c, pat="OUT_OF_FRAME"),"landWt"])/sum(df2$landWt)*100,4)
```

 # produces final objects
  cs4s2013<-df1
  cs4s2014<-df2

 # compiles stratification options into an object
  StratificationOpt<-list(original.file=file.info("cs4_all_checked_Shake_correct.rds"), stratification_date = Sys.time(), psuType=psuType,
  psuPT = psuPT, psuES=psuES, psuBC=psuBC, psuOT=psuOT, psuALL=psuALL,
  min_psu_per_strata=min_psu_per_strata, n_ssu_per_psu=n_ssu_per_psu,
  n_main_strata_regional=n_main_strata_regional,
n_main_strata_national_pt=n_main_strata_national_pt,n_main_strata_national_es=n_main_strata_national_es,
n_main_strata_national_bc=n_main_strata_national_bc,
  allocType=allocType, perc_land_regional=perc_land_regional, perc_land_regional_pt=perc_land_regional_pt,
perc_land_regional_es=perc_land_regional_es, perc_land_regional_bc=perc_land_regional_bc)

 # final QCA (compares input matrix and output matrix)
  dim(db1)[1]==dim(cs4s2013)[1]+dim(cs4s2014)[1]
  tapply(db1$landWt, db1$onShoreSampCtr, sum) == tapply(cs4s2013$landWt, cs4s2013$onShoreSampCtr,
sum)+tapply(cs4s2014$landWt, cs4s2014$onShoreSampCtr, sum)

 # saves data and settings
  save (cs4s2013, cs4s2014, StratificationOpt, file=paste("cs4_Shake_AllVessels_prepared_for_sim_",paste(format(Sys.time(),
"%Y%m%d%H%M"), sep=""),".Rdata", sep=""))

cs4_Shake_All_Vessels_Final_Data_Preparation_For_Sims.r

#============================
fishPi - WP2 - CS2 - SHake - Final Data Preparation For Simulation - 'All Vessels' Setting
#============================

 # load packages
 library(data.table)
 library(sampling)
 library(survey)
 library(stringr)

# load data
 load("data_input\\cs4_Shake_AllVessels_prepared_for_sim_201512061735.Rdata")

Note on Scenarions
 # "NONE"   # no stratification
 # "Port2a"   # stratification by Port with effort allocation proportional to landings
 # "Quarter2b" # stratification by Quarter with effort allocation proportional to landings (all quarters selected)
 # "PortQuarter2c" # stratification by Port*Quarter with effort allocation proportional to landings
 # "COUNTRY"   # stratification by Ctr/Institute with current effort (scenario extra)
 # "Ctr30"   # stratification by Ctr/Institute with effort allocation proportional to landings
 # "Port3a"   # stratification by Ctr/Institute and Port (within Ctr/Institute) with effort allocation proportional to landings
 # "Quarter3b" # stratification by Ctr/Institute and Quarter (within Ctr/Institute) with effort allocation proportional to landings (all
quarters selected)
 # "PortQuarter3c" # stratification by Ctr/Institute and Port*Quarter (within Ctr/Institute) with effort allocation proportional to landings

 for (Strata in c("Port2a","Quarter2b","PortQuarter2c"))
 {

 # ==
 # Preparation of sim dataset
  # includes all days possible in all strata, completing the matrix
  # creates pop_study (the study population) from pop (the population)
 # ==

 print(Strata)
 print("preparing...")
 ptc <- Sys.time()

 dados <- cs4s2014

 dados$landWt <- dados$landWt/1000

 names(dados)[names(dados)==Strata] <- "Stratum"

```
 df1 <- as.data.table(dados);

 df1 <- df1[order(Stratum, onShoreSampLoc, arvDate)]

 df1 <- df1[,list(landWt = sum(landWt)), by=.(stratum = Stratum, quarter  = onShoreSampQuarter, market = onShoreSampLoc, market_day
= paste(onShoreSampLoc, arvDate), market_day_event = paste(onShoreSampLoc, arvDate, voyageId), arvDate = arvDate)]


 # adds days without trips
  if(Strata=="Port2a"){
   a <- expand.grid(sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("market","arvDate")
  }
  if(Strata=="Quarter2b"){
   a <- expand.grid(sort(unique(dados$onShoreSampQuarter)), sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("quarter", "market","arvDate")
   # delete impossible combinations
   a <- a[a$quarter == quarters(a$arvDate),]
  }
  if(Strata=="PortQuarter2c"){
   a <- expand.grid(sort(unique(dados$onShoreSampQuarter)), sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("quarter", "market","arvDate")
   # delete impossible combinations
   a <- a[a$quarter == quarters(a$arvDate),]
  }

 # creates market_day
  a$market_day <- paste(a$market, a$arvDate)

 # adds stratification info
  if(Strata=="Port2a"){
   tab_consulta <- unique(df1[,.(market, stratum)])
   a$stratum <- tab_consulta$stratum[match(a$market,tab_consulta$market)]
  }
  if(Strata=="Quarter2b"){
   tab_consulta <- unique(df1[,.(quarter, stratum)])
   a$stratum <- tab_consulta$stratum[match(a$quarter,tab_consulta$quarter)]
  }
  if(Strata=="PortQuarter2c"){
   tab_consulta <- unique(df1[,.(market_day, market, quarter, stratum)])
   a$stratum <- tab_consulta$stratum[match(paste(a$market_day, a$quarter),paste(tab_consulta$market_day,tab_consulta$quarter))]
   a[is.na(a$stratum),]$stratum <- tab_consulta$stratum[match(paste(a[is.na(a$stratum),]$market,
a[is.na(a$stratum),]$quarter),paste(tab_consulta$market,tab_consulta$quarter))]
   a[is.na(a$stratum),]$stratum <- tab_consulta$stratum[match(a[is.na(a$stratum),]$market,tab_consulta$market)]
  }

 # adds market_day_event info e landWt where it exists
   b <- merge(a, df1[,.(market_day, market_day_event, landWt)], by="market_day", all.x = T)

 # fills in info for days without trips
   if(Strata=="Port2a"){
    b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
   }
   if(Strata=="Quarter2b"){
    b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"quarter"],
b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
    b[is.na(b$stratum),]$stratum <- as.character(b[is.na(b$stratum),]$quarter)
   }
   if(Strata=="PortQuarter2c"){
    b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"quarter"],
b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
   }

   b[is.na(b$landWt),"landWt"] <- 0
   if(sum(names(b)=="quarter")==0)
   {
   b$quarter <- quarters(b$arvDate)
   }

   df2 <- as.data.table(b)
```

```
 # adds weekdays
  df2[, `:=`(weekday,format(arvDate, format="%u"))]

   # weekday analyses: about 9% in saturday and sunday
   df2[, sum(landWt)/sum(df2$landWt), by = weekday]

 # adds numerical strata and psus
  df2[, c("stratum1","market_day1","market_day_event1"):= list(as.numeric(as.factor(stratum)),as.numeric(as.factor(market_day)),
as.numeric(as.factor(market_day_event)))];

 # orders
  df2 <- df2[order(stratum1,market_day1, market_day_event1),]

 # pop is the total population, i.e., including OUT_OF_FRAME
  pop <- df2

 # pop_study is the population effectively sampled, i.e., excluding the OUT_OF_FRAME
  pop_study <- df2[!grepl(stratum, pat = "OUT_OF_FRAME"),]

  # coverage of pop_study (in market_events)
  dim(pop_study)[1]/dim(pop)[1]


# ============================================
# imports sampling effort
# ============================================

 if(Strata=="Port2a") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc2a_samples.csv", sep=";", header = T)}
 if(Strata=="Quarter2b") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc2b_samples.csv", sep=";", header = T)}
 if(Strata=="PortQuarter2c") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc2c_samples.csv", sep=";", header = T)}


  names(samp_dist)[names(samp_dist)==Strata] <- "stratum"

  key_stratum1 <- unique(pop[,list(stratum, stratum1),])

  samp_dist$stratum1 <- key_stratum1$stratum1[match(as.character(samp_dist$stratum),key_stratum1$stratum,)]
   # reordenacao de colunas
    samp_dist <- samp_dist[c("stratum","stratum1","n_psu","n_ssu_per_psu")]
   # alteracao do nome de colunas
    names(samp_dist)[names(samp_dist)=="n_psu"] <- "n1" # = n_psu
    names(samp_dist)[names(samp_dist)=="n_ssu_per_psu"] <- "n2" # = n_ssu

# adds market_day and market_day_event to dados
 dados$market_day <- paste(dados$onShoreSampLoc, dados$arvDate)
 dados$market_day_event <- paste(dados$onShoreSampLoc, dados$arvDate, dados$voyageId)


print("preparation done!")


# ============================================
# saves outputs
# ============================================

 print("saving....")
 output_file <- paste("Data_cs4_AllVessels_",Strata,"_prepared_",format(Sys.time(), "%Y%m%d%H%M"),".Rdata", sep="")
 save(dados, pop, pop_study, samp_dist,  file = paste("Data_prepared\\",output_file, sep=""))
 write(output_file, file = paste("Data_prepared\\Data_cs4_AllVessels_",Strata,"_last_prepared.txt",sep=""))
 print("saving done!")
 print(Sys.time()-ptc)


}


for (Strata in c("NONE","COUNTRY","Ctr30", "Port3a","Quarter3b","PortQuarter3c"))
{

# ============================================
```

```
# Preparation of sim dataset
 # includes all days possible in all strata, completing the matrix
 # creates pop_study (the study population) from pop (the population)
# ===========================================

print(Strata)
print("preparing...")
ptc <- Sys.time()

 dados <- cs4s2014

 dados$landWt <- dados$landWt/1000

 if(Strata=="NONE"){ # this bit uses the Port3a as a guide in order to plug in the OUT_OF_FRAME
  names(dados)[names(dados)=="Port3a"] <- "Stratum"
  dados$Stratum[!grepl(dados$Stratum, pat="OUT_OF_FRAME")] <- "NONE"
  }
 if(Strata=="COUNTRY"){
  dados$Stratum <- str_sub(dados$onShoreSampLoc,1,2);
  dados[!dados$Stratum %in% c("BC","ES","PT"),"Stratum"] <- "OUT_OF_FRAME"
  # adds the OUT_OF_FRAME that are inside the national strata
  dados[grepl(dados$Port3a, pat="OUT_OF_FRAME"),"Stratum"] <- dados[grepl(dados$Port3a, pat="OUT_OF_FRAME"),"Port3a"]
  }
 if(!Strata %in% c("COUNTRY","NONE")){
  names(dados)[names(dados)==Strata] <- "Stratum"
  }

  df1 <- as.data.table(dados);
  df1 <- df1[order(Stratum, onShoreSampLoc, arvDate)]
  df1 <- df1[,list(landWt = sum(landWt)), by=.(stratum = Stratum, country = onShoreSampCtr, quarter = onShoreSampQuarter, market =
onShoreSampLoc, market_day = paste(onShoreSampLoc, arvDate), market_day_event = paste(onShoreSampLoc, arvDate, voyageId),
arvDate = arvDate)]


 # adds market*days without market_events
  if(Strata=="NONE" | Strata=="COUNTRY"){
   a <- expand.grid(sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("market","arvDate")
   }
   if(Strata=="Ctr30"){
   a <- expand.grid(sort(unique(dados$onShoreSampCtr)), sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("country", "market","arvDate")
   # deletes impossible combinations
   a <- a[!(a$country=="BC" & str_sub(a$market,1,2)!="BC"),]
   a <- a[!(a$country=="ES" & str_sub(a$market,1,2)!="ES"),]
   a <- a[!(a$country=="PT" & str_sub(a$market,1,2)!="PT"),]
  a <- a[!(a$country=="OT" & str_sub(a$market,1,2) %in% c("PT","BC","ES")),]
  }
  if(Strata=="Port3a"){
   a <- expand.grid(sort(unique(dados$onShoreSampCtr)), sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("country", "market","arvDate")
   # deletes impossible combinations
   a <- a[!(a$country=="BC" & str_sub(a$market,1,2)!="BC"),]
   a <- a[!(a$country=="ES" & str_sub(a$market,1,2)!="ES"),]
   a <- a[!(a$country=="PT" & str_sub(a$market,1,2)!="PT"),]
  a <- a[!(a$country=="OT" & str_sub(a$market,1,2) %in% c("PT","BC","ES")),]
  }
  if(Strata=="Quarter3b"){
   a <- expand.grid(sort(unique(dados$onShoreSampCtr)), sort(unique(dados$onShoreSampQuarter)),
sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
   names(a) <- c("country","quarter", "market","arvDate")
   # deletes impossible combinations
   a <- a[a$quarter == quarters(a$arvDate),]
   a <- a[!(a$country=="BC" & str_sub(a$market,1,2)!="BC"),]
   a <- a[!(a$country=="ES" & str_sub(a$market,1,2)!="ES"),]
   a <- a[!(a$country=="PT" & str_sub(a$market,1,2)!="PT"),]
   a <- a[!(a$country=="OT" & str_sub(a$market,1,2) %in% c("PT","BC","ES")),]
  }
  if(Strata=="PortQuarter3c"){
   a <- expand.grid(sort(unique(dados$onShoreSampCtr)), sort(unique(dados$onShoreSampQuarter)),
sort(unique(dados$onShoreSampLoc)), sort(unique(dados$arvDate)))
```

```
    names(a) <- c("country","quarter", "market","arvDate")
    # deletes impossible combinations
    a <- a[a$quarter==quarters(a$arvDate),]
    a <- a[!(a$country=="BC" & str_sub(a$market,1,2)!="BC"),]
    a <- a[!(a$country=="ES" & str_sub(a$market,1,2)!="ES"),]
    a <- a[!(a$country=="PT" & str_sub(a$market,1,2)!="PT"),]
    a <- a[!(a$country=="OT" & str_sub(a$market,1,2) %in% c("PT","BC","ES")),]
   }


   # creates market_day field
    a$market_day <- paste(a$market, a$arvDate)
   # adds stratum info
    if(Strata=="NONE"){
     tab_consulta <- unique(df1[,.(market, stratum)])
     a$stratum <- tab_consulta$stratum[match(a$market,tab_consulta$market)]
    }
    if(Strata=="COUNTRY"){
     tab_consulta <- unique(df1[,.(market, stratum)])
     a$stratum <- tab_consulta$stratum[match(a$market,tab_consulta$market)]
    }
    if(Strata=="Ctr30"){
     tab_consulta <- unique(df1[,.(market, stratum)])
     a$stratum <- tab_consulta$stratum[match(a$market,tab_consulta$market)]
    }
    if(Strata=="Port3a"){
     tab_consulta <- unique(df1[,.(market, stratum)])
     a$stratum <- tab_consulta$stratum[match(a$market,tab_consulta$market)]
    }
    if(Strata=="Quarter3b"){
     tab_consulta <- unique(df1[,.(country, quarter, stratum)])
     a$stratum <- tab_consulta$stratum[match(paste(a$country,a$quarter),paste(tab_consulta$country,tab_consulta$quarter))]
     a[is.na(a$stratum),"stratum"] <- "OT OUT_OF_FRAME"
    }
    if(Strata=="PortQuarter3c"){
     tab_consulta <- unique(df1[,.(country, market_day, market, quarter, stratum)])
     a$stratum <- tab_consulta$stratum[match(paste(a$market_day, a$quarter),paste(tab_consulta$market_day,tab_consulta$quarter))]
     a[is.na(a$stratum),]$stratum <- tab_consulta$stratum[match(paste(a[is.na(a$stratum),]$market,
a[is.na(a$stratum),]$quarter),paste(tab_consulta$market,tab_consulta$quarter))]
     a[is.na(a$stratum),]$stratum <- tab_consulta$stratum[match(a[is.na(a$stratum),]$market,tab_consulta$market)]
    }

    # adds market_day_event info and landWt where it exists
     b <- merge(a, df1[,.(market_day, market_day_event, landWt)], by="market_day", all.x = T)

    # fills in info of market_days with no market_day_events
     if(Strata=="NONE"){
      b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
     }
     if(Strata=="COUNTRY"){
      b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
     }
     if(Strata=="Ctr30"){
      b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
     }
     if(Strata=="Port3a"){
      b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
     }
     if(Strata=="Quarter3b"){
      b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"quarter"],
b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
      b[is.na(b$stratum),]$stratum <- as.character(b[is.na(b$stratum),]$quarter)
     }
     if(Strata=="PortQuarter3c"){
      b[is.na(b$landWt),"market_day_event"] <- paste(b[is.na(b$landWt),"quarter"],
b[is.na(b$landWt),"market"],b[is.na(b$landWt),"arvDate"],"NONE")
     }

     b[is.na(b$landWt),"landWt"] <- 0

     if(sum(names(b)=="quarter")==0)
```

```
    {
    b$quarter <- quarters(b$arvDate)
    }


   df2 <- as.data.table(b)


 # creates weekday
  df2[, `:=`(weekday,format(arvDate, format="%u"))]

   # analise: cerca de 9% ao sabado e domingo
   df2[, sum(landWt)/sum(df2$landWt), by = weekday]

 # creats numerical strata and psus
  df2[, c("stratum1","market_day1","market_day_event1"):= list(as.numeric(as.factor(stratum)),as.numeric(as.factor(market_day)),
as.numeric(as.factor(market_day_event)))];

 # sorts
  df2 <- df2[order(stratum1,market_day1, market_day_event1),]


 # pop is the total population, i.e., including OUT_OF_FRAME
  pop <- df2
 # pop_study is the population effectively sampled, i.e., excluding the OUT_OF_FRAME
  pop_study <- df2[!grepl(stratum, pat = "OUT_OF_FRAME"),]

  # cobertura de pop_study (em viagens)
  dim(pop_study)[1]/dim(pop)[1]


# ============================================
# imports sampling effort
# ============================================


 if(Strata=="NONE") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_ScN_samples.csv", sep=";", header = T)}  # preparado a mao
 if(Strata=="COUNTRY") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc30p_samples.csv", sep=";", header = T)} # preparado a
mao
 if(Strata=="Ctr30") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc30_samples.csv", sep=";", header = T)}  # decorrente de
script de alocacao de esforco
 if(Strata=="Port3a") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc3a_samples.csv", sep=";", header = T)}  # decorrente de
script de alocacao de esforco
 if(Strata=="Quarter3b") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc3b_samples.csv", sep=";", header = T)}  # decorrente
de script de alocacao de esforco
 if(Strata=="PortQuarter3c") {samp_dist <- read.table("samp_dist\\Shake_AllVessels_Sc3c_samples.csv", sep=";", header = T)}  #
decorrente de script de alocacao de esforco


  names(samp_dist)[names(samp_dist)==Strata] <- "stratum"

  key_stratum1 <- unique(pop[,list(stratum, stratum1),])

  samp_dist$stratum1 <- key_stratum1$stratum1[match(as.character(samp_dist$stratum),key_stratum1$stratum,)]
   # reordenacao de colunas
    samp_dist <- samp_dist[c("stratum","stratum1","n_psu","n_ssu_per_psu")]
   # alteracao do nome de colunas
    names(samp_dist)[names(samp_dist)=="n_psu"] <- "n1" # = n_psu
    names(samp_dist)[names(samp_dist)=="n_ssu_per_psu"] <- "n2" # = n_ssu


# adds market_day and market_day_event to dados
 dados$market_day <- paste(dados$onShoreSampLoc, dados$arvDate)
 dados$market_day_event <- paste(dados$onShoreSampLoc, dados$arvDate, dados$voyageId)


print("preparation done!")


# ============================================
# saves outputs
```

```
# ============================================

  print("saving....")
  output_file <- paste("Data_cs4_AllVessels_",Strata,"_prepared_",format(Sys.time(), "%Y%m%d%H%M"),".Rdata", sep="")
  save(dados, pop, pop_study, samp_dist,  file = paste("Data_prepared\\",output_file, sep=""))
  write(output_file, file = paste("Data_prepared\\Data_cs4_AllVessels_",Strata,"_last_prepared.txt",sep=""))
  print("saving done!")
  print(Sys.time()-ptc)


}
```

cs4_Shake_All_Vessels_Simulations.r

```
#============================
# fishPi - WP2 - CS2 - SHake - Simulation code - 'All Vessels' Setting
#============================

# Notes on Scenarions
 # "NONE"    # no stratification
 # "Port2a"   # stratification by Port with effort allocation proportional to landings
 # "Quarter2b"  # stratification by Quarter with effort allocation proportional to landings (all quarters selected)
 # "PortQuarter2c" # stratification by Port*Quarter with effort allocation proportional to landings
 # "COUNTRY"    # stratification by Ctr/Institute with current effort (scenario extra)
 # "Ctr30"    # stratification by Ctr/Institute with effort allocation proportional to landings
 # "Port3a"   # stratification by Ctr/Institute and Port (within Ctr/Institute) with effort allocation proportional to landings
 # "Quarter3b"  # stratification by Ctr/Institute and Quarter (within Ctr/Institute) with effort allocation proportional to landings (all
quarters selected)
 # "PortQuarter3c" # stratification by Ctr/Institute and Port*Quarter (within Ctr/Institute) with effort allocation proportional to landings


  for (Strata in c("NONE","Port2a","Quarter2b","PortQuarter2c","COUNTRY","Ctr30","Port3a","Quarter3b","PortQuarter3c"))
  {

 # ============================================
 # Sample selection and estimation
 # ============================================

  # loads packages
   require(data.table)
   require(beepr)
   require(snowfall)  # parallel computing
   source("cs4_Shake_All_Vessels_Sim_Func.R") # simulation functions

  # parameters [only "srswr" allowed, "srswor" still being developed]
   samp1 <- "srswr"
   samp2 <- "srswr"
   imprimir.QCA <- FALSE
   nsim<-1000


  print(paste("Strata=",Strata))
  ptc1<-Sys.time()

  # loads data
  last_file<-scan(paste("data_prepared\\Data_cs4_AllVessels_",Strata,"_last_prepared.txt", sep=""), what="raw", quiet=TRUE);
print(last_file)
  load(paste("data_prepared\\",last_file, sep=""))

  # starts parallel computing
  sfInit(parallel=TRUE, cpus=4)
  sfLibrary(data.table)
  sfExport("nsim", "dados", "samp1", "samp2", "imprimir.QCA","samp_dist", "pop_study",
"pop","FAZ_STRATIFIED_TWO_STAGE_CLUSTER_SAMPLING_AND_ESTIMATION","prn")

  # starts the seeds
  sfLapply(1:nsim, prn)

  # runs simulations
  res<-sfLapply(1:nsim, FAZ_STRATIFIED_TWO_STAGE_CLUSTER_SAMPLING_AND_ESTIMATION)
```

```
 # stops parallel computing
 sfStop()

 # saves outputs
  print("saving....")
  output_file<-paste("Data_cs4_AllVessels_",Strata,"_res_",format(Sys.time(), "%Y%m%d%H%M"),".Rdata", sep="")
  save(res, file=paste("results\\",output_file, sep=""))
  write(output_file, file=paste("results\\Data_cs4_AllVessels_",Strata,"_last_res.txt",sep=""))

 print(paste("first phase:",round(difftime(Sys.time(), ptc1, units="sec"),2), "secs")); beep("ping")

 # some time to cool down
 #Sys.sleep(60)


# ============================================
# Domain estimation and result compilation
# ============================================


 # loads data and results
 last_file<-scan(paste("data_prepared\\Data_cs4_AllVessels_",Strata,"_last_prepared.txt", sep=""), what="raw", quiet=TRUE);
print(last_file)
 load(paste("data_prepared\\",last_file, sep=""))

 last_file<-scan(paste("results\\Data_cs4_AllVessels_",Strata,"_last_res.txt", sep=""), what="raw", quiet=TRUE); print(last_file)
 load(paste("results\\",last_file, sep=""))

 ptc2<-Sys.time()

 # additional data preparation
  dados<-as.data.table(dados)
  dados[dados$rect==""]$rect<-"99u99"
  dados[,c("rect"):= toupper(rect),]
  setkey(dados, market_day_event)
  pop_study<-as.data.table(pop_study)

 # starts parallel computing
  sfInit(parallel=TRUE, cpus=4)
  sfLibrary(data.table)
  sfExport("res","dados","pop_study","pop", "FAZ_STRATIFIED_TWO_STAGE_CLUSTER_RESULTS","FAZ_DISTRIB_CONTRIB")

 # runs domain estimation
  resultados_parc<-sfLapply(res, FAZ_STRATIFIED_TWO_STAGE_CLUSTER_RESULTS)

 # stops parallel computing
  sfStop()

 # compiles results
  resultados<-FAZ_STRATIFIED_TWO_STAGE_CLUSTER_RESULT_COMPILATION(resultados_parc, nsim, Strata, samp1, samp2)

 print(paste("second phase:",round(difftime(Sys.time(), ptc2, units="sec"),2), "secs")); beep("ping")

 # saves outputs
  print("saving....")
  output_file<-paste("Data_cs4_AllVessels_",Strata,"_results_",format(Sys.time(), "%Y%m%d%H%M"),".Rdata", sep="")
  save(resultados, nsim, samp1, samp2, Strata, file=paste("results\\",output_file, sep=""))
  write(output_file, file=paste("results\\Data_cs4_AllVessels_",Strata,"_last_results.txt",sep=""))

 a<-difftime(Sys.time(), ptc2, units="sec")
 print(paste("total time:",round(a,2), "for", nsim, "simulations, i.e.,", round(a/nsim,2),"secs per run"))

 #Sys.sleep(60)

 beep("fanfare")
}
```

cs4_Shake_All_Vessels_Sim_Func.R

```
prn <- function(seed){
```

```
 # function to set seed in multiple cores (More details in Jochen Knaus, Christine Porzelius Parallel computing using R package snowfall)
 set.seed(seed)

  #demonstracao
  #sfLapply(1:10, prn)
  #b<-sfLapply(1:10, function(x){rnorm(2)});b
 }


FAZ_STRATIFIED_TWO_STAGE_CLUSTER_SAMPLING_AND_ESTIMATION <- function(lo)
{

#===================================================================================
# fishPi - WP2 - CS2 - SHake - Function for Two Stage Cluster Sampling and Estimation
#===================================================================================



 samp_dist1 <- samp_dist
 pop_study1 <- pop_study
 pop1 <- pop


 # ============================================
 # Sample Selection and Probabilities
 # ============================================

  # function that corrects "sample" execution in the case of only one element in the population (more info: ?sample and ?sample.int)
  sample2 <- function(x, ...) x[sample.int(length(x), ...)]


  # psu part

   list_psu <- pop_study[, list(market_day1 = unique(market_day1)), by = stratum1]

   # adds sampling effort

    list_psu <- merge(list_psu, samp_dist1, by="stratum1", all.x = T)
    list_psu <- merge(list_psu,list_psu[,list(N1=.N), by = stratum1], by="stratum1", all.x = T)

    list_psu <- list_psu[order(list_psu$stratum1),]

   # probabilities psu

    if(samp1=="srswr")
     {
     list_psu$prob1 <- 1-(1-1/list_psu$N1)**list_psu$n1

      # detects situations when sample.int is used em issues error (means that the script needs to be improved)
      if(sum(list_psu$N1==1)>0){stop(paste("estrato com N=1 => necessario adaptar script com sample2"))}

      sample_psu <- list_psu[, melt(Map(sample, split(market_day1, stratum1), size=unique(list_psu[,list(stratum1, n1)])$n1,
MoreArgs=list(replace=TRUE)))]

     }

     if(samp1=="srswor") # sem reposicao
     {
     stop ("srswor not developed")
     }

    names(sample_psu)<-c("market_day1","stratum1")

  # ssu part

   list_ssu <- pop_study[pop_study$market_day1 %in% sample_psu$market_day1,.(stratum1,market_day1, market_day_event1) ]


   if(samp1=="srswr")
    {
     a <- table(sample_psu$market_day1)
     list_ssu$replicates <- a[match(list_ssu$market_day1,as.numeric(names(a)))]
```

```
    } else {list_ssu$replicates <- 1}

   # adds ssu to be sampled e evaluates prob2
    list_ssu <- merge(list_ssu, samp_dist, by="stratum1", all.x = T)
    list_ssu <- merge(list_ssu,list_ssu[,list(N2=.N), by = market_day1], by="market_day1", all.x = T, sort = FALSE)

    list_ssu <- list_ssu[order(list_ssu$market_day1, list_ssu$market_day_event1),]

    if(samp1=="srswr")
     {
      list_ssu$N2_old <- list_ssu$N2
      list_ssu$N2 <- list_ssu$N2*list_ssu$replicates
      list_ssu$n2 <- list_ssu$n2*list_ssu$replicates
     }


    if(samp2=="srswr") # com reposicao
     {
     list_ssu$prob2 <- 1-(1-1/list_ssu$N2_old)**(list_ssu$n2/list_ssu$replicates)

     list_ssu1 <- list_ssu[N2_old==1,] # situations that must be handled with sample2
     list_ssu2 <- list_ssu[!N2_old==1,] # situations that can be handled with sample


     if(dim(list_ssu1)[1]>0) # situations that must be handled with sample2
      {

      sample_ssu <- rbind(
          list_ssu1[, melt(Map(sample2, Map(rep, split(list_ssu1$market_day_event1,
list_ssu1$market_day1),unique(list_ssu1[,list(market_day1, replicates)])$replicates), size = unique(list_ssu1[,list(market_day1, n2)])$n2,
MoreArgs = list(replace = TRUE)))],
          list_ssu2[, melt(Map(sample, Map(rep, split(list_ssu2$market_day_event1,
list_ssu2$market_day1),unique(list_ssu2[,list(market_day1, replicates)])$replicates), size = unique(list_ssu2[,list(market_day1, n2)])$n2,
MoreArgs = list(replace = TRUE)))])



      } else { # situations that can be handled with sample

      sample_ssu <- list_ssu[, melt(Map(sample, split(market_day_event1, market_day1), size = unique(list_ssu[,list(market_day1, n2)])$n2,
MoreArgs = list(replace = TRUE)))]

      }

     }



    if(samp2=="srswor") # sem reposicao
     {
     stop ("srswor not developed")
     }


    names(sample_ssu)<-c("market_day_event1","market_day1")
    sample_ssu$market_day1 <- as.numeric(sample_ssu$market_day1)

    sample_ssu$stratum1 <- list_ssu$stratum1[match(sample_ssu$market_day1, list_ssu$market_day1)]

    sample_ssu$prob1 <- list_psu$prob1[match(sample_ssu$market_day1, list_psu$market_day1)]
    sample_ssu$prob2 <- list_ssu$prob2[match(sample_ssu$market_day_event1, list_ssu$market_day_event1)]
    sample_ssu$N1 <- list_psu$N1[match(sample_ssu$market_day1, list_psu$market_day1)]
    sample_ssu$n1 <- list_psu$n1[match(sample_ssu$market_day1, list_psu$market_day1)]
    sample_ssu$N2 <- list_ssu$N2_old[match(sample_ssu$market_day_event1, list_ssu$market_day_event1)]
    sample_ssu$n2 <- list_ssu$n2[match(sample_ssu$market_day_event1, list_ssu$market_day_event1)] /
list_ssu$replicates[match(sample_ssu$market_day_event1, list_ssu$market_day_event1)]

    sample_ssu <- sample_ssu[order(sample_ssu$stratum1,
sample_ssu$market_day1),c("stratum1","market_day1","market_day_event1","N1","n1","N2","n2", "prob1","prob2")]
```

```
  my.sample <- merge(sample_ssu,pop1[,.(market_day,quarter,
market,arvDate,stratum,market_day_event,landWt,weekday,market_day_event1)], by="market_day_event1",all.x = T)

  # handling of duplicated psu under srswr (1= saida 1, 2= saida 2)
  my.sample$saida <- 1
  my.sample <- as.data.table(my.sample)
  lista <- my.sample[, .N>samp_dist$n2[1],by = market_day1][V1==TRUE]$market_day1
  for (i in lista)
   {
   a <- my.sample[market_day1==i,]
   nrep <- dim(a)[1]/a$n2[1]
   my.sample[market_day1==i,]$saida <- as.numeric(rep(1:nrep,samp_dist$n2[1]))
   }
  my.sample <- as.data.frame(my.sample)


  # QCA
  if(imprimir.QCA==TRUE)
   {
   print("unique(market_day1)")
   print(tapply(my.sample$market_day1, my.sample$stratum1, function(x){length(unique(x))}))
   print("no de viagens1")
   print(tapply(my.sample$market_day1, my.sample$stratum1, function(x){length(x)}))
   print("objectivos de amostragem no market_days")
   print(tapply(samp_dist$n1, samp_dist$stratum1, sum))
   print("objectivos de amostragem no viagens")
   print(tapply(samp_dist$n1*samp_dist$n2, samp_dist$stratum1, sum))
   }


  # joint probability
  my.sample$prob <- my.sample$prob1*my.sample$prob2

  # sample weight
  my.sample$pw <- 1/my.sample$prob



 # ============================================
 # Estimation
 # ============================================

  stratum_totest <- stratum_varest <- matrix(nrow = length(unique(pop1$stratum1)), ncol = 2)
  stratum_totest[,1]<-unique(pop1$stratum1)
  stratum_varest[,1]<-unique(pop1$stratum1)

  # totals [str_est is auxiliary]
  my.sample <- as.data.table(my.sample)
  ssu_est <- my.sample[,list(variancia = var(landWt), media = mean(landWt), m2=.N, n1 = unique(n1)), by = list(stratum1,
market_day1,saida, N1, N2)]
  psu_est <- ssu_est[, list(psu_tot = N2*media, m2), by = list(stratum1, market_day1,saida, N1)]
  str_est <- psu_est[, list(str_tot = N1*mean(psu_tot)), by = list(stratum1, N1)]

  stratum_totest[,2]<-str_est$str_tot[match(stratum_totest[,1], str_est$stratum1)]
  tot_study <- sum(str_est$str_tot)
  tot_pop <- dim(unique(pop1[,list(stratum1, market_day1)]))[1]/sum(str_est$N1)*sum(str_est$str_tot)

  # variance [str_est is auxiliary]
  ssu_est <- my.sample[,list(ssu_var = var(landWt), m2=.N, n1 = unique(n1)), by =  list(stratum1, market_day1, saida, N1, M2 = N2)]
  ssu_est <- ssu_est[, list(ssu_parc = M2*(M2-m2)*ssu_var/m2, m2 = m2, M2 = M2), by = list(stratum1, market_day1, saida)]
  str_part1 <- ssu_est[, list(ssu_temp = sum(ssu_parc), n = sum(m2)), by = list(stratum1)]
  str_part1$N <- list_psu$N[match(str_part1$stratum1, list_psu$stratum1)]
  str_part1$str_part1 <- str_part1$N/str_part1$n*str_part1$ssu_temp

  str_part2 <- psu_est[, list(psu_temp = var(psu_tot), n = sum(m2)), by = list(stratum1)]
  str_part2$N <- list_psu$N[match(str_part2$stratum1, list_psu$stratum1)]
  str_part2$str_part2 <- str_part2$N/str_part2$n*(str_part2$N-str_part2$n)*str_part2$psu_temp

  str_est <- merge(str_part1[,list(stratum1, n, N, str_part1)], str_part2[,list(stratum1, str_part2)], by="stratum1", all.x = T)
  str_est$str_tot <- str_part1$str_part1+str_part2$str_part2
```

```
  stratum_varest[,2]<-str_est$str_tot[match(stratum_varest[,1], str_est$stratum1)]

  var_study <- sum(str_est$str_tot)
  var_pop <- dim(unique(pop1[,list(stratum1, market_day1)]))[1]/sum(str_est$N)*sum(str_est$str_tot)

  cv_study <- round(sqrt(var_study)/tot_study*100,2)
  cv_pop <- round(sqrt(var_pop)/tot_pop*100,2)


 # ====================
 # Output
 # ====================

  estimacao <- list(tot_study = tot_study,tot_pop = tot_pop, var_study = var_study, var_pop = var_pop, cv_study = cv_study, cv_pop = cv_pop, stratum_totest = stratum_totest, stratum_varest = stratum_varest, sample_psu = sample_psu, sample_ssu = sample_ssu)


 estimacao

}


FAZ_DISTRIB_CONTRIB <- function(variavel, dados2, ls1){

 #=====================================================================================
 # fishPi - WP2 - CS2 - SHake - Aux Function for Domain Estimation
 #=====================================================================================


 temp <- dcast(dados2, market_day_event~temp_var, fun.aggregate = sum, value.var="landWt")
  # aux variable
   nfact <- nlevels(factor(dados2$temp_var))-1
  # adds sample_ssu to dcast
   x <- as.data.table(ls1$sample_ssu)
   setkey(x,market_day_event)
   a <- as.data.frame(temp[J(x[,list(market_day_event,ssu_contrib,landWt)]),])
  # handles 0s
   for(j in (ncol(a)-2)-(nfact:0))
   {
   a[is.na(a[,j]), j]<-0}
  # totals sampled within variable
   b1 <- apply(a[(ncol(a)-2)-(nfact:0)],2,sum, na.rm = T)
  # number of positive trips within variable
   b2 <- apply(a[(ncol(a)-2)-(nfact:0)]>0,2,sum, na.rm = T)
  #evaluates proportion within variable
   for(j in (ncol(a)-2)-(nfact:0))
    {
    a[, j]<-a[, j]/a$landWt*a$ssu_contrib
    a[, j][is.na(a[, j])]<-0
    }
 list(a, b1, b2)
  }


FAZ_STRATIFIED_TWO_STAGE_CLUSTER_RESULTS <- function(x)
{

 #=====================================================================================
 # fishPi - WP2 - CS2 - SHake - Function for compilation of main results and domain estimation
 #=====================================================================================


 df1 <- pop_study[, list(landWt_tot = sum(landWt)), by = market_day1]
 tot_study <- tot_pop <- var_study <- var_pop <- cv_study <- cv_pop <- c()
 n_duplicates_psu <- n_duplicates_ssu <- n_psu_with_0_hke <- n_ssu_with_0_hke <- c()
 n_psu_by_market <- n_psu_by_quarter <- n_psu_by_weekday <- n_psu_by_marketquarter <- c()
 n_ssu_by_market <- n_ssu_by_quarter <- n_ssu_by_weekday <- n_ssu_by_marketquarter <- c()
 n_ssu_hke_by_quarter <- n_ssu_hke_by_market <- n_ssu_hke_by_marketquarter <-  c()
 ton_sampled_ssu_by_quarter <- ton_sampled_ssu_by_market <- ton_sampled_ssu_by_marketquarter <- c()

 ton_hke_total_in_psu <- c()
```

```
  ton_hke_in_ssu <- c()

  tot_study_by_quarter <- tot_study_by_market <- tot_study_by_marketquarter <- c()
  tot_study_vslFlgCtry <- tot_study_rect <- tot_study_foCatEu6 <- tot_study_area <- tot_study_fleetNat <- tot_study_domainNat1 <-
tot_study_onShoreStratNat <- tot_study_onShoreSampCtr <- c()
  n_ssu_with_hke_vslFlgCtry <- n_ssu_with_hke_rect <- n_ssu_with_hke_foCatEu6 <- n_ssu_with_hke_area <-
n_ssu_with_hke_domainNat1 <- n_ssu_with_hke_fleetNat <- n_ssu_with_hke_onShoreStratNat <- n_ssu_with_hke_onShoreSampCtr <-
c()
  ton_hke_sampled_vslFlgCtry <- ton_hke_sampled_rect <- ton_hke_sampled_foCatEu6 <- ton_hke_sampled_area <-
ton_hke_sampled_domainNat1 <- ton_hke_sampled_fleetNat <- ton_hke_sampled_onShoreStratNat <-
ton_hke_sampled_onShoreSampCtr <- c()

  stratum_totest <- matrix(nrow = length(unique(pop$stratum1)), ncol = 2)
  stratum_totest[,1]<-unique(pop$stratum1)
  stratum_varest <- matrix(nrow = length(unique(pop$stratum1)), ncol = 2)
  stratum_varest[,1]<-unique(pop$stratum1)

  ptc <- Sys.time();

  ls1 <- x

  tot_study <- ls1[[1]]
  tot_pop <- ls1[[2]]
  var_study <- ls1[[3]]
  var_pop <- ls1[[4]]
  cv_study <- ls1[[5]]
  cv_pop <- ls1[[6]]
  stratum_totest[,2]<-ls1[[7]][,2]
  stratum_varest[,2]<-ls1[[8]][,2]
  n_duplicates_psu <- sum(duplicated(ls1$sample_psu$market_day))
  n_duplicates_ssu <- sum(duplicated(ls1$sample_ssu$market_day_event1))

  # adds market to output
   ls1$sample_psu$market <- pop_study$market[match(ls1$sample_psu$market_day1, pop_study$market_day1)]
   ls1$sample_ssu$market <- pop_study$market[match(ls1$sample_ssu$market_day1, pop_study$market_day1)]
  # adds quarter to output
   ls1$sample_psu$quarter <- pop_study$quarter[match(ls1$sample_psu$market_day1, pop_study$market_day1)]
   ls1$sample_ssu$quarter <- pop_study$quarter[match(ls1$sample_ssu$market_day1, pop_study$market_day1)]
  # adds weekday to output
   ls1$sample_psu$weekday <- pop_study$weekday[match(ls1$sample_psu$market_day1, pop_study$market_day1)]
   ls1$sample_ssu$weekday <- pop_study$weekday[match(ls1$sample_ssu$market_day1, pop_study$market_day1)]
  # adds landWt to output
   ls1$sample_psu$landWt_tot <- df1$landWt[match(ls1$sample_psu$market_day1, df1$market_day1)]
   ls1$sample_ssu$landWt <- pop_study$landWt[match(ls1$sample_ssu$market_day_event1, pop_study$market_day_event1)]
  # adds market_day_event to output
   ls1$sample_ssu$market_day_event <- pop_study$market_day_event[match(ls1$sample_ssu$market_day_event1,
pop_study$market_day_event1)]

  # ====================
  # domain-estimation
  # ====================

  # n_psus sampled
   n_psu_by_market <- table(ls1$sample_psu$market)
   n_psu_by_quarter <- table(ls1$sample_psu$quarter)
   n_psu_by_weekday <- table(ls1$sample_psu$weekday)
   n_psu_by_marketquarter <- melt(table(ls1$sample_psu$market, ls1$sample_psu$quarter))
   n_psu_by_marketquarter$ID <- paste(n_psu_by_marketquarter$Var1, n_psu_by_marketquarter$Var2)

  # n_ssus sampled
   n_ssu_by_market <- table(ls1$sample_ssu$market)
   n_ssu_by_quarter <- table(ls1$sample_ssu$quarter)
   n_ssu_by_weekday <- table(ls1$sample_ssu$weekday)
   n_ssu_by_marketquarter <- melt(table(ls1$sample_ssu$market, ls1$sample_ssu$quarter))
   n_ssu_by_marketquarter$ID <- paste(n_ssu_by_marketquarter$Var1, n_ssu_by_marketquarter$Var2)

  # ton sampled in ssu by quarter, market and marketquarter
   ton_sampled_ssu_by_quarter <- tapply(ls1$sample_ssu$landWt, ls1$sample_ssu$quarter, sum)
   ton_sampled_ssu_by_market <- tapply(ls1$sample_ssu$landWt, ls1$sample_ssu$market, sum)
   ton_sampled_ssu_by_marketquarter <- melt(tapply(ls1$sample_ssu$landWt, list(ls1$sample_ssu$market, ls1$sample_ssu$quarter),
sum))
```

```
  ton_sampled_ssu_by_marketquarter$ID <- paste(ton_sampled_ssu_by_marketquarter$Var1,
ton_sampled_ssu_by_marketquarter$Var2)

 # n ssu with hke by quarter, market and marketquarter
  n_ssu_hke_by_quarter <- tapply(ls1$sample_psu$landWt>0, ls1$sample_psu$quarter, sum)
  n_ssu_hke_by_market <- tapply(ls1$sample_psu$landWt>0, ls1$sample_psu$market, sum)
  n_ssu_hke_by_marketquarter <- melt(tapply(ls1$sample_psu$landWt>0, list(ls1$sample_psu$market, ls1$sample_psu$quarter), sum))
  n_ssu_hke_by_marketquarter$ID <- paste(n_ssu_hke_by_marketquarter$Var1, n_ssu_hke_by_marketquarter$Var2)

 # contributions at fleet level (using weights)
  ls1$sample_ssu$ssu_contrib <-
ls1$sample_ssu$landWt/((ls1$sample_ssu$n1/ls1$sample_ssu$N1)*(ls1$sample_ssu$n2/ls1$sample_ssu$N2))

 # totals by quarter, market and marketquarter
  tot_study_by_quarter <- tapply(ls1$sample_ssu$ssu_contrib, ls1$sample_ssu$quarter, sum)
  tot_study_by_market <- tapply(ls1$sample_ssu$ssu_contrib, ls1$sample_ssu$market, sum)
  tot_study_by_marketquarter <- melt(tapply(ls1$sample_ssu$ssu_contrib, list(ls1$sample_ssu$market, ls1$sample_ssu$quarter), sum))
  tot_study_by_marketquarter$ID <- paste(tot_study_by_marketquarter$Var1, tot_study_by_marketquarter$Var2)

 # totals (other)

  dados[,c("temp_var") := eval(parse(text="vslFlgCtry"))]
  b <- FAZ_DISTRIB_CONTRIB(variavel="vslFlgCtry", dados2 = dados, ls1)
  tot_study_vslFlgCtry <- apply(b[[1]][,levels(factor(dados$temp_var))],2, sum)
  ton_hke_sampled_vslFlgCtry <- b[[2]]
  n_ssu_with_hke_vslFlgCtry <- b[[3]]

  dados[,c("temp_var") := eval(parse(text="rect"))]
  b <- FAZ_DISTRIB_CONTRIB(variavel="rect", dados2 = dados, ls1)
  tot_study_rect <- apply(b[[1]][,levels(factor(dados$temp_var))],2, sum)
  ton_hke_sampled_rect <- b[[2]]
  n_ssu_with_hke_rect <- b[[3]]

  dados[,c("temp_var") := eval(parse(text="foCatEu6"))]
  b <- FAZ_DISTRIB_CONTRIB(variavel="foCatEu6", dados2 = dados, ls1)
  tot_study_foCatEu6 <- apply(b[[1]][,levels(factor(dados$temp_var))],2, sum)
  ton_hke_sampled_foCatEu6 <- b[[2]]
  n_ssu_with_hke_foCatEu6 <- b[[3]]

  dados[,c("temp_var") := eval(parse(text="area"))]
  b <- FAZ_DISTRIB_CONTRIB(variavel="area", dados2 = dados, ls1)
  tot_study_area <- apply(b[[1]][,levels(factor(dados$temp_var))],2, sum)
  ton_hke_sampled_area <- b[[2]]
  n_ssu_with_hke_area <- b[[3]]

  dados[,c("temp_var") := eval(parse(text="fleetNat"))]
  b <- FAZ_DISTRIB_CONTRIB(variavel="fleetNat", dados2 = dados, ls1)
  tot_study_fleetNat <- apply(b[[1]][,levels(factor(dados$temp_var))],2, sum)
  ton_hke_sampled_fleetNat <- b[[2]]
  n_ssu_with_hke_fleetNat <- b[[3]]

  dados[,c("temp_var") := eval(parse(text="domainNat1"))]
  b <- FAZ_DISTRIB_CONTRIB(variavel="domainNat1", dados2 = dados, ls1)
  tot_study_domainNat1 <- apply(b[[1]][,levels(factor(dados$temp_var))],2, sum)
  ton_hke_sampled_domainNat1 <- b[[2]]
  n_ssu_with_hke_domainNat1 <- b[[3]]


 # "efficiency"
  n_psu_with_0_hke <- sum(ls1$sample_psu$landWt==0)
  n_ssu_with_0_hke <- sum(ls1$sample_ssu$landWt==0)

  ton_hke_total_in_psu <- sum(ls1$sample_psu$landWt)
  ton_hke_in_ssu <- sum(ls1$sample_ssu$landWt)

 print(Sys.time()-ptc)

 # ===================
 # Output
 # ===================
```

```
  analises <- list(n_duplicates_psu = n_duplicates_psu, n_duplicates_ssu = n_duplicates_ssu, n_psu_with_0_hke = n_psu_with_0_hke,
n_ssu_with_0_hke = n_ssu_with_0_hke, ton_hke_total_in_psu = ton_hke_total_in_psu, ton_hke_in_ssu = ton_hke_in_ssu,
     n_psu_by_market = n_psu_by_market, n_psu_by_quarter = n_psu_by_quarter, n_psu_by_weekday =
n_psu_by_weekday,n_psu_by_marketquarter = n_psu_by_marketquarter,
      n_ssu_by_market = n_ssu_by_market, n_ssu_by_quarter = n_ssu_by_quarter, n_ssu_by_weekday =
n_ssu_by_weekday,n_ssu_by_marketquarter = n_ssu_by_marketquarter,
       n_ssu_hke_by_quarter = n_ssu_hke_by_quarter, n_ssu_hke_by_market = n_ssu_hke_by_market, n_ssu_hke_by_marketquarter =
n_ssu_hke_by_marketquarter,
        ton_sampled_ssu_by_quarter = ton_sampled_ssu_by_quarter, ton_sampled_ssu_by_market = ton_sampled_ssu_by_market,
ton_sampled_ssu_by_marketquarter = ton_sampled_ssu_by_marketquarter,
        tot_study_by_quarter = tot_study_by_quarter, tot_study_by_market = tot_study_by_market, tot_study_by_marketquarter =
tot_study_by_marketquarter,
        tot_study_vslFlgCtry  =  tot_study_vslFlgCtry,
         ton_hke_sampled_vslFlgCtry = ton_hke_sampled_vslFlgCtry,
         n_ssu_with_hke_vslFlgCtry = n_ssu_with_hke_vslFlgCtry,
        tot_study_rect = tot_study_rect,
         ton_hke_sampled_rect = ton_hke_sampled_rect,
         n_ssu_with_hke_rect = n_ssu_with_hke_rect,
        tot_study_foCatEu6 = tot_study_foCatEu6,
         ton_hke_sampled_foCatEu6 = ton_hke_sampled_foCatEu6,
         n_ssu_with_hke_foCatEu6 = n_ssu_with_hke_foCatEu6,
        tot_study_area = tot_study_area,
         ton_hke_sampled_area = ton_hke_sampled_area,
         n_ssu_with_hke_area = n_ssu_with_hke_area,
        tot_study_fleetNat = tot_study_fleetNat,
         ton_hke_sampled_fleetNat = ton_hke_sampled_fleetNat,
         n_ssu_with_hke_fleetNat = n_ssu_with_hke_fleetNat ,
        tot_study_onShoreStratNat = tot_study_onShoreStratNat,
         ton_hke_sampled_onShoreStratNat = ton_hke_sampled_onShoreStratNat,
         n_ssu_with_hke_onShoreStratNat = n_ssu_with_hke_onShoreStratNat,
        tot_study_onShoreSampCtr = tot_study_onShoreSampCtr,
         ton_hke_sampled_onShoreSampCtr = ton_hke_sampled_onShoreSampCtr,
         n_ssu_with_hke_onShoreSampCtr = n_ssu_with_hke_onShoreSampCtr,
        tot_study_domainNat1 = tot_study_domainNat1,
         ton_hke_sampled_domainNat1 = ton_hke_sampled_domainNat1,
         n_ssu_with_hke_domainNat1 = n_ssu_with_hke_domainNat1)

  resultados <- list(tot_study = tot_study,tot_pop = tot_pop, var_study = var_study, var_pop = var_pop, cv_study = cv_study, cv_pop =
cv_pop, stratum_totest = stratum_totest, stratum_varest = stratum_varest, sample_psu = ls1$sample_psu, sample_ssu = ls1$sample_ssu,
      analises = analises)

  resultados


}

FAZ_STRATIFIED_TWO_STAGE_CLUSTER_RESULT_COMPILATION <- function(resultados_parc, nsim, Strata, samp1, samp2)
{

 #==================================================================================
 # fishPi - WP2 - CS2 - SHake - Function for final compilation of results
 #==================================================================================


 resultados <- list()


  for (i in 1:nsim)
  {
   resultados$tot_study <- c(resultados$tot_study,resultados_parc[[i]][["tot_study"]])
   resultados$tot_pop <- c(resultados$tot_pop,resultados_parc[[i]][["tot_pop"]])
   resultados$var_study <- c(resultados$var_study,resultados_parc[[i]][["var_study"]])
   resultados$var_pop <- c(resultados$var_pop,resultados_parc[[i]][["var_pop"]])
   resultados$cv_study <- c(resultados$cv_study,resultados_parc[[i]][["cv_study"]])
   resultados$cv_pop <- c(resultados$cv_pop,resultados_parc[[i]][["cv_pop"]])
   if(i==1){
    resultados$stratum_totest <- resultados_parc[[i]][["stratum_totest"]]
    resultados$stratum_varest <- resultados_parc[[i]][["stratum_varest"]]
     } else {
    resultados$stratum_totest <- cbind(resultados$stratum_totest,resultados_parc[[i]][["stratum_totest"]][,2])
    resultados$stratum_varest <- cbind(resultados$stratum_varest,resultados_parc[[i]][["stratum_varest"]][,2])
```

```
    }
   # duplicates in psu and ssu
   resultados$analises$n_duplicates_psu <- 
c(resultados$analises$n_duplicates_psu,resultados_parc[[i]][["analises"]][["n_duplicates_psu"]])
   resultados$analises$n_duplicates_ssu <- 
c(resultados$analises$n_duplicates_ssu,resultados_parc[[i]][["analises"]][["n_duplicates_ssu"]])

   # psu and ssu with 0 hake
   resultados$analises$n_psu_with_0_hke <- 
c(resultados$analises$n_psu_with_0_hke,resultados_parc[[i]][["analises"]][["n_psu_with_0_hke"]])
   resultados$analises$n_ssu_with_0_hke <- 
c(resultados$analises$n_ssu_with_0_hke,resultados_parc[[i]][["analises"]][["n_ssu_with_0_hke"]])

   # ton of hke in psu (total) and ssu (sampled)
   resultados$analises$ton_hke_total_in_psu <- 
c(resultados$analises$ton_hke_total_in_psu,resultados_parc[[i]][["analises"]][["ton_hke_total_in_psu"]])
   resultados$analises$ton_hke_in_ssu <- c(resultados$analises$ton_hke_in_ssu,resultados_parc[[i]][["analises"]][["ton_hke_in_ssu"]])

   # psus by market, quarter, weekday and marketquarter
    resultados$analises$n_psu_by_market <- 
cbind(resultados$analises$n_psu_by_market,resultados_parc[[i]][["analises"]][["n_psu_by_market"]])
    resultados$analises$n_psu_by_quarter <- 
cbind(resultados$analises$n_psu_by_quarter,resultados_parc[[i]][["analises"]][["n_psu_by_quarter"]])
    resultados$analises$n_psu_by_weekday <- 
cbind(resultados$analises$n_psu_by_weekday,resultados_parc[[i]][["analises"]][["n_psu_by_weekday"]])
    a <- resultados_parc[[i]][["analises"]][["n_psu_by_marketquarter"]][c("ID","value")]
    resultados$analises$n_psu_by_marketquarter <- cbind(resultados$analises$n_psu_by_marketquarter,tapply(a$value, a$ID, sum))

   # ssus by market, quarter, weekday and marketquarter
    resultados$analises$n_ssu_by_market <- 
cbind(resultados$analises$n_ssu_by_market,resultados_parc[[i]][["analises"]][["n_ssu_by_market"]])
    resultados$analises$n_ssu_by_quarter <- 
cbind(resultados$analises$n_ssu_by_quarter,resultados_parc[[i]][["analises"]][["n_ssu_by_quarter"]])
    resultados$analises$n_ssu_by_weekday <- 
cbind(resultados$analises$n_ssu_by_weekday,resultados_parc[[i]][["analises"]][["n_ssu_by_weekday"]])
    a <- resultados_parc[[i]][["analises"]][["n_ssu_by_marketquarter"]][c("ID","value")]
    resultados$analises$n_ssu_by_marketquarter <- cbind(resultados$analises$n_ssu_by_marketquarter,tapply(a$value, a$ID, sum))

   # n ssus with hke by market, quarter and marketquarter
    resultados$analises$n_ssu_hke_by_market <- 
cbind(resultados$analises$n_ssu_hke_by_market,resultados_parc[[i]][["analises"]][["n_ssu_hke_by_market"]])
    resultados$analises$n_ssu_hke_by_quarter <- cbind(resultados$analises$n_ssu_hke_by_quarter, 
resultados_parc[[i]][["analises"]][["n_ssu_hke_by_quarter"]])
    # check
    a <- resultados_parc[[i]][["analises"]][["n_ssu_hke_by_marketquarter"]][c("ID","value")]
    resultados$analises$n_ssu_hke_by_marketquarter <-  cbind(resultados$analises$n_ssu_hke_by_marketquarter,tapply(a$value, a$ID, 
sum))

   # ton hke sampled in ssus per market, quarter and marketquarter
   resultados$analises$ton_sampled_ssu_by_market <- cbind(resultados$analises$ton_sampled_ssu_by_market, 
resultados_parc[[i]][["analises"]][["ton_sampled_ssu_by_market"]])
   resultados$analises$ton_sampled_ssu_by_quarter <- cbind(resultados$analises$ton_sampled_ssu_by_quarter, 
resultados_parc[[i]][["analises"]][["ton_sampled_ssu_by_quarter"]])
   a <- resultados_parc[[i]][["analises"]][["ton_sampled_ssu_by_marketquarter"]][c("ID","value")]
   resultados$analises$ton_sampled_ssu_by_marketquarter <- 
cbind(resultados$analises$ton_sampled_ssu_by_marketquarter,tapply(a$value, a$ID, sum))

   resultados$analises$tot_study_by_quarter <- 
cbind(resultados$analises$tot_study_by_quarter,resultados_parc[[i]][["analises"]][["tot_study_by_quarter"]])
   resultados$analises$tot_study_by_market <- 
cbind(resultados$analises$tot_study_by_market,resultados_parc[[i]][["analises"]][["tot_study_by_market"]])
   a <- resultados_parc[[i]][["analises"]][["tot_study_by_marketquarter"]][c("ID","value")]
   resultados$analises$tot_study_by_marketquarter <- cbind(resultados$analises$tot_study_by_marketquarter,tapply(a$value, a$ID, 
sum))

   # outputs of domain estimation
    if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_vslFlgCtry"]]))
    {
     resultados$analises$tot_study_vslFlgCtry <- 
cbind(resultados$analises$tot_study_vslFlgCtry,resultados_parc[[i]][["analises"]][["tot_study_vslFlgCtry"]])
     resultados$analises$ton_hke_sampled_vslFlgCtry <-
```

```
cbind(resultados$analises$ton_hke_sampled_vslFlgCtry,resultados_parc[[i]][["analises"]][["ton_hke_sampled_vslFlgCtry"]])
    resultados$analises$n_ssu_with_hke_vslFlgCtry <-
cbind(resultados$analises$n_ssu_with_hke_vslFlgCtry,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_vslFlgCtry"]])
   }
   if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_rect"]]))
   {
    resultados$analises$tot_study_rect <- cbind(resultados$analises$tot_study_rect,resultados_parc[[i]][["analises"]][["tot_study_rect"]])
    resultados$analises$ton_hke_sampled_rect <-
cbind(resultados$analises$ton_hke_sampled_rect,resultados_parc[[i]][["analises"]][["ton_hke_sampled_rect"]])
    resultados$analises$n_ssu_with_hke_rect <-
cbind(resultados$analises$n_ssu_with_hke_rect,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_rect"]])
   }
   if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_foCatEu6"]]))
   {
    resultados$analises$tot_study_foCatEu6 <-
cbind(resultados$analises$tot_study_foCatEu6,resultados_parc[[i]][["analises"]][["tot_study_foCatEu6"]])
    resultados$analises$ton_hke_sampled_foCatEu6 <-
cbind(resultados$analises$ton_hke_sampled_foCatEu6,resultados_parc[[i]][["analises"]][["ton_hke_sampled_foCatEu6"]])
    resultados$analises$n_ssu_with_hke_foCatEu6 <-
cbind(resultados$analises$n_ssu_with_hke_foCatEu6,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_foCatEu6"]])
   }
   if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_area"]]))
   {
    resultados$analises$tot_study_area <-
cbind(resultados$analises$tot_study_area,resultados_parc[[i]][["analises"]][["tot_study_area"]])
    resultados$analises$ton_hke_sampled_area <-
cbind(resultados$analises$ton_hke_sampled_area,resultados_parc[[i]][["analises"]][["ton_hke_sampled_area"]])
    resultados$analises$n_ssu_with_hke_area <-
cbind(resultados$analises$n_ssu_with_hke_area,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_area"]])
   }
   if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_fleetNat"]]))
   {
    resultados$analises$tot_study_fleetNat <-
cbind(resultados$analises$tot_study_fleetNat,resultados_parc[[i]][["analises"]][["tot_study_fleetNat"]])
    resultados$analises$ton_hke_sampled_fleetNat <-
cbind(resultados$analises$ton_hke_sampled_fleetNat,resultados_parc[[i]][["analises"]][["ton_hke_sampled_fleetNat"]])
    resultados$analises$n_ssu_with_hke_fleetNat <-
cbind(resultados$analises$n_ssu_with_hke_fleetNat,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_fleetNat"]])
   }
   if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_onShoreStratNat"]]))
   {
    resultados$analises$tot_study_onShoreStratNat <-
cbind(resultados$analises$tot_study_onShoreStratNat,resultados_parc[[i]][["analises"]][["tot_study_onShoreStratNat"]])
    resultados$analises$ton_hke_sampled_onShoreStratNat <-
cbind(resultados$analises$ton_hke_sampled_onShoreStratNat,resultados_parc[[i]][["analises"]][["ton_hke_sampled_onShoreStratNat"]])
    resultados$analises$n_ssu_with_hke_onShoreStratNat <-
cbind(resultados$analises$n_ssu_with_hke_onShoreStratNat,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_onShoreStratNat"]])
    }
   if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_onShoreSampCtr"]]))
   {
    resultados$analises$tot_study_onShoreSampCtr <-
cbind(resultados$analises$tot_study_onShoreSampCtr,resultados_parc[[i]][["analises"]][["tot_study_onShoreSampCtr"]])
    resultados$analises$ton_hke_sampled_onShoreSampCtr <-
cbind(resultados$analises$ton_hke_sampled_onShoreSampCtr,resultados_parc[[i]][["analises"]][["ton_hke_sampled_onShoreSampCtr"]])
    resultados$analises$n_ssu_with_hke_onShoreSampCtr <-
cbind(resultados$analises$n_ssu_with_hke_onShoreSampCtr,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_onShoreSampCtr"]])
   }
    if(!is.null(resultados_parc[[i]][["analises"]][["tot_study_domainNat1"]]))
   {
    resultados$analises$tot_study_domainNat1 <-
cbind(resultados$analises$tot_study_domainNat1,resultados_parc[[i]][["analises"]][["tot_study_domainNat1"]])
    resultados$analises$ton_hke_sampled_domainNat1 <-
cbind(resultados$analises$ton_hke_sampled_domainNat1,resultados_parc[[i]][["analises"]][["ton_hke_sampled_domainNat1"]])
    resultados$analises$n_ssu_with_hke_domainNat1 <-
cbind(resultados$analises$n_ssu_with_hke_domainNat1,resultados_parc[[i]][["analises"]][["n_ssu_with_hke_domainNat1"]])
    }

  }

  resultados$nsim <- nsim
  resultados$strata <- Strata
```

```
  resultados$samp1 <- samp1
  resultados$samp2 <- samp2

  resultados

}
```

ANNEX 15 WP3 A regional sampling plan for the collection of PETS data

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP3 – Regional sampling programmes for fisheries and ecosystem impact data not currently collected (including stomach contents and bycatch)

Deliverable 3.1 - A regional sampling plan for data collection of PETS (Protected, Endangered, and Threatened Species)

Lead: AZTI, Spain

Core Team: Estanis Mugerza, Lucia Zarauz and Eider Andonegi (AZTI, Spain)
Hugo Mendes and Ana Moreno (IPMA, Portugal)
Mike Armstrong, (CEFAS, UK)
Patrik Börjesson (SLU, Sweden)

Table of Contents

1 Framework

Interactions between fisheries and non-target species such as protected, endangered and threatened species (PETS), including cetaceans, sea birds, turtles some elasmobranchs and rare fish species, are frequent and widespread leading to levels of incidental mortality which, in some cases, pose a serious threat to certain species or populations. Such interactions can also have an adverse effect on fishing productivity and profitability.

Under the current Data Collection Framework (Council regulation (EC) No. 199/2008), there are no binding obligations for Member States (MS) to collect data on species other than commercial fish and certain invertebrates. This means that the current DCF does not fulfil the requirements of agreed and prioritised RSC (Regional Sea Convection) indicators on the number of drowned mammals and water birds in fishing gears (HELCOM) and the numbers of individual mammals within species being by-caught in relation to population (agreed as OSPAR indicator in the North Sea).

However, under several EU instruments (Habitats Directive 92/43/EEC, Council Regulation (EC) No. 812/2004 and Birds Directive 2009/147/EC) MS are required to monitor and report on bycatch of certain non-fisheries protected species, including cetaceans, sea birds and marine turtles.

There remain, however, important gaps in knowledge about such incidental catches, including their scale, spatial distribution, importance in different fisheries and the effectiveness of mitigation measures.

Regional bycatch monitoring programmes will enable quantification of the effects of fishing operations on bycatch species and, as such, have an important role to play in the development of ecosystem based approach management. These data can also provide useful biological information on PETS, and technical information on aspects of gear use that can inform management decisions.

2 Regional sampling programme data collection scheme (stages in design and implementation)

Regional cooperation and task sharing have been strongly promoted within the present Data Collection Framework (DCF). The main way to support a regional approach has been through the Regional Coordination Meetings (RCMs), established by the Commission. So far these meetings have been successful in collating national meta data on the performance of fisheries and sampling to a regional level. This has led to a more common understanding of fisheries and sampling within a region. It is foreseeable that the emphasis of a regional approach in data collection will be even stronger in the revised DCF since the reform of the Common Fisheries Policy will include greater regionalisation of fisheries management. Stocks and fisheries are further assessed at a regional/international level. Regional data collection programmes need to be documented and based on statistically sound survey methods to allow for quality control and transparency in the data collection assessment advice process.

The different steps in designing and implementing a regional data collection scheme to meet end-user needs are illustrated in Figure 1. The most critical first stages are for the end users to clearly define the objectives and estimates required at a regional level to support fisheries management or conservation objectives, and an indication of minimum precision needed (steps 2 & 3). For example, for a data-rich stock of high economic or conservation importance, this could be estimates of stock status (biomass and fishing mortality) with sufficient precision to allow detailed estimation of Maximum Sustainable Yield (MSY) reference points, stock status and Total Allowable Catch (TAC) forecast options. For a relatively minor, data-limited stock, simpler metrics may be required (e.g. a series of survey abundance indices and fishery landings). The subsequent steps 4 -6 (type of data required; data collection methods and design, sampling intensity and allocation of sampling effort across countries and strata) cannot take place in an effective way without the information specified in step 3. The relationship between the cost of data collection and the precision achieved should be considered when specifying a regional catch sampling programme, although it may take some time to develop the data required for such optimisation and until then a judgment-based approach may be required.

The subsequent steps are the data collection, the handling of the data (e.g. archiving in the regional data bases), evaluation of data quality (quality indicators) and finally the analysis of the data to provide the required estimates and associated measures of uncertainty. The evaluation of sampling schemes against benchmarks for good practice, and the monitoring of data quality using suitable indicators, should be given considerable attention. The data analysis (step 9) may provide additional evidence of problems with data quality that may be traced back to sampling schemes, leading to improvements in sampling design.

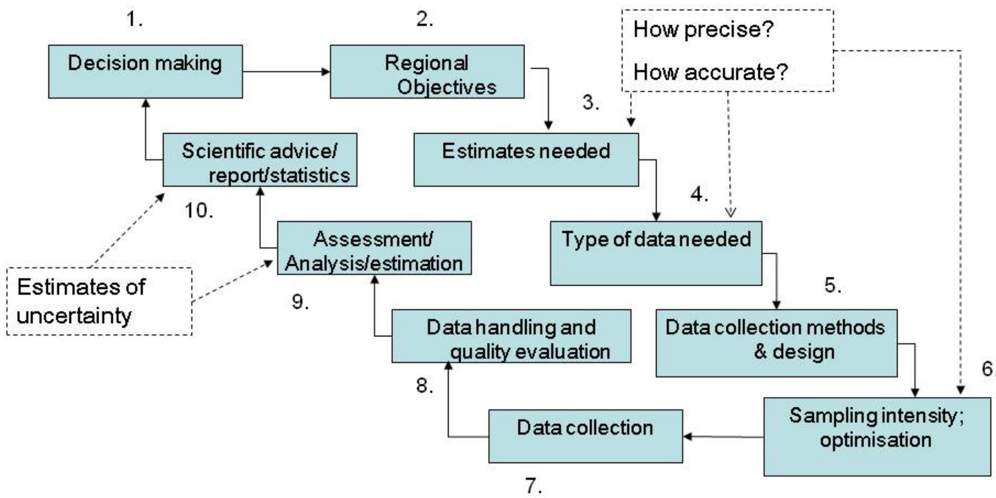


Figure1. Regional sampling programme scheme.

The above mentioned regional sampling scheme and stages should be also adopted for the data collection on PETS. In the following section these specific stages will be covered focused on PETS.

2.1 Decision making

Decision making is the first stage in a regional sampling programme and relevant end-users need to be identified.

End users are defined in the DCF Regulation as bodies with a research or management interest in the scientific analysis of data in the fisheries sector (Article 2 of the DCF Regulation). For the purposes of defining regional sampling plans, the end users can be considered to be advisory bodies such as ICES, STECF or RFMOs who require a variety of data and estimates on a recurrent basis at a range of aggregation, from detailed (anonymised) data to highly aggregated data, to develop advice on fisheries management. At national level end users can be a variety of national governments, local bodies, NGO's and others, all with a clear interest and legitimate requirement for fisheries and marine ecosystem data.

In Table 1 the most relevant end-users of PETS bycatch issues have been identify. The data exploitation of these end users is also considered.

Table 1. Relevant end-users.

End User	End user sub groups	Use of by catch data
ICES	WGBYC (Working Group on By-Catch of Protected Species)	Collates and assesses information on bycatch monitoring and assessment for protected species, including mammals, birds, turtles, and rare fish.
	WGMME (Working Group on Marine Mammal Ecology)	Provides scientific advice in relation to marine mammals. examines any new information on population sizes, population/stock structure and management frameworks for marine mammals and assess how these can contribute to the regulatory requirements
	JWGBIRD (Joint OSPAR/HELCOM/ICES Working Group on Seabirds)	Requests for advice from OSPAR that, recently have featured the development of Ecology Quality Objectives (EcoQOs) and the development of common bird indicators under the EU's Marine Strategy Framework Directive
	ICES expert assessment WG	
	WGFTFB (Working Group on Fishing Technology and Fish Behaviour)	Studies measurements and observations relating to scientific and commercial fishing gears, design and statistical methods and operations, and fish behaviour in relation to fishing.
	ICES expert groups and steering groups dealing with integrated ecosystem assessment	
	WGCATCH (Working Group on Commercial Catches)	Documents national fishery sampling schemes, establishes best practice and guidelines on sampling and estimation procedures, and provides advice on other uses of fishery data.
	WGRFS (Working Group on Recreational Fisheries Surveys)	Planning and coordination of marine recreational fishery data collection for stock assessment purposes.
Other RMFO (ICCAT, NAFO, NEAFC…)	Expert assessment and ecosystem WG	
European Commission	DGMARE & DG Environment	Implementation of MSFD; achievement of GES with good management of recreational as well as commercial fishery impacts. Implementation of CE 812/2004, Birds Directive, Habitats Directive
	STECF	Inclusion of data collection in the EU MAP
International Organizations	FAO, OSPAR, ASCOBANS, ACAP, IWC, HELCOM	Identifying threats, recommending action plans, implementation of different agreements.

		For MSFD purposes The regional Sea conventions are one of the absolutely most important end user since they're the ones working with the indicators/indicator targets for D1 which are later implemented at national level.
Regional Coordination Groups	RCGs for each region	Coordination and cost-effectiveness of by catch data collection within regions (if included in EU-MAP)
National Governments and regional fisheries authorities within countries		Developing policy positions on management that reflects the ecosystem aspects of sustainable development in coastal regions and spatial planning such as MCZs. Meeting international agreed responsibilities
Scientific community in general.	Universities; Govt. departments; other Institutes	Scientists interested on by catch and ecosystem dynamics Data for publication
Representative bodies for International and national commercial fisheries.	Commercial fishermen's organisations and federations.	Policy developments in relation to interaction between commercial species and main predators;
Recreational fisheries bodies	Recreational fishermen's organisations and federations (EAA, Angling Trust…)	Developing best practices
Advisory Councils	e.g. North Western Waters AC; North Sea AC…..	Policy developments in relation to interaction between commercial species and main predators;
Marine NGOs	Birdlife international, WWF, GREENPEACE, OCEANA etc.	Policy developments in relation to interaction between commercial species and main predators;

2.2 Regional objectives and Estimates

To define the regional objectives and estimates on the data to be collected it is necessary to have a clear list of PETS species, but at present no such list exists. All the species which are protected under National law and under Union legislation (Habitats Directive 92/43/EEC, Birds Directive 2009/147/EC) international agreements and National law should be included. This includes all cetaceans, seals, sea birds and marine turtles species. There are some fish species that have the same protection, such as Shads, (*Alosa sps*.), lampreys (*Lampetra fluviatilis* and *Petromyzon marinus)* and the sturgeon (*Acipenser sturio*). All these species should be monitored and covered by these regional sampling programmes.

There are other species (e.g. some sharks and ray species) where their conservation status is considered at risk by the International Union for Conservation Nature (IUCN) or other conservation authorities. These species should be considered and analysed on a case by case basis at regional level.

In order to facilitate the inclusion of these species in the regional plans and the process of monitoring the incidental bycatch, an updated list of species relevant at regional level should be compiled. This list will be flexible and updated over time in coordination between the main end users and the Regional Co-ordination Groups (RCGs).

RCGs will be the responsible bodies for defining regional sampling objectives. To this aim, needs of the end users (estimates, aggregation and precision levels) the feasibility of collecting such information and their potential impact on data collection programmes (available resources) will need to be taken into account.

2.3 Type of data needed

ICES WGBYC and Scientific, Technical and Economic Committee for Fisheries (STECF) as main end-users have worked on the type of data needed and practical issues that need to be taken into account (STECF 14-02 and STECF 14-07). The type of data outlined below, could be considered as basic general data. However, it is possible that for some species, fisheries or regions, extra data or higher resolution data will be needed. This task should be carried out among the main end-users and the different RCGs doing a case by case data needs analysis. This work could be done intersessionally in specific subgroups of experts in bycatch (i.e. members from WGBYC), experts in sampling (i.e. members from WGCATCH) and RCGs.

These are the descriptions of the types of data required and some practical issues that need to be taken into account according to some of the main end users (STECF 14-02 and STECF 14-07):

Vessel ID/Date/Time/Haul ID/% of operation sampled
This data is similar to the data that is collected in discard sampling programmes, but the approach for incidental bycatch is very different. Discards consists by definition of a large volume of small specimens, comparable to the target catch. This part of the catch can be sampled by taking a small subsample (basket). For incidental bycatch this is not possible and it needs to be recorded at the haul level. This means inspection of the opening of the cod-end or a scan of the catch during handling. As hauls are concurrently sampled for discards and retained catch, it is important that the sampling protocols contain a checkbox indicating that the haul was actually checked for incidental bycatch and in the case of a scan during hauling, an indicator of the percentage coverage. This process is important as it enables the identification of hauls or sets with zero bycatch.

Geographical position
In general, geographic position should be expressed in latitude/longitude (degrees and minutes). If the exact location is not known or available, the approximate location should be fitted to the geographical area/grid in use by ICES. (Rectangle, subdivision, division, geographical subarea).

Gear
As a minimum, FAO gear codes or DCF level 4 codes should be used. If possible, DCF level 6 or further disaggregation is recommended in order to link the sampling with the fisheries characterization.
Mesh size/number of hooks and type

The mesh size of gillnets and trammel nets is of interest as it influences the likelihood of entanglement.

The number and type of longline hooks will also influence the probability of bycatch.

Species
If it is not possible to identify to species level, it should be recorded on a higher taxonomic level (group of species, genus, and family or order level). This is of particular importance for the recording of seabirds which includes a large number of possible species for many areas. Protocols should include a list of rare species that should be recorded during trips. These species should have a code in the institute database and code lists should be available to the person who enters the data in the database. It has been recognized that most countries do not have codes for a lot of protected, endangered and threatened species, which causes data not to be stored in national databases (ICES 2013a). Comprehensive species lists and entry codes are provided by the AFIS List of Species for Fisheries Statistics Purposes by WoRMS (World Register of Marine Species). The use of these internationally agreed lists is recommended, because they allow the merging of data coming from different countries and therefore are essential to set up a regionally shared data base.

http://www.fao.org/fishery/collection/asfis/en.
http://www.marinespecies.orgrms

Number of specimens

Number of specimens by species.

Indicator of dead or alive

Yes or No.

Mitigation type

Sampling should contain information on any mitigation measures applied. Currently so called Acoustic Deterrent Devices are obligatory in some fisheries under Council Regulation (EC) No. 812/2004. Brand, type and indicators of adequate use should be collected as well. Other mitigation measures (i.e. for turtles, bird) may come into use in the future.

2.4 Data collection methods and sampling design

Different methodologies are used to collect data on PETS. Although direct observations are the preferred means of estimating bycatch rates, these are sometimes impractical, usually because they are expensive or because space on small scale fleet limits the acceptance of observers onboard (ICES 2011c).

In this report both indirect and direct methodologies will be analysed. The strengths and weakness should be also considered taking into account different criteria (e.g. cost, data quality etc.).

2.4.1 Indirect observation methodologies

Indirect observations methodologies could be used mainly to obtain a general overview on bycatch events. These methodologies are low cost. On the other hand lower resolution and quality data is obtained.

Anecdotal Accounts

Anecdotal accounts of marine mammal, seabird and turtle bycatches in fisheries may provide some initial evidence that bycatch occur in an area. These accounts are usually not random as news of exceptional rather than common events are more likely to be spread. The information usually is not detailed and may be biased. The information collected in this way may increase awareness of the potential for high bycatch risk in a fishery in a certain area which may then lead to more specific monitoring needs.

Stranding/floating

The presence of dead animals, with evidence of gear interactions, on coasts or at-sea may highlight the fact that some bycatch is occurring in a region. As a quantitative measure such observations are not of much use because the number of dead animals that wash ashore is not necessarily directly related to the animals that are by-caught in any given region.

Care must be taken not to over-interpret data from stranded animals, and protocols for establishing cause of death must be followed. Strandings can help augment other data sources and raise awareness of bycatch in an area.

Photo identification-identification studies

Studies of scars, injuries and gear remains on cetaceans, seals, birds and turtles resulting from fisheries interactions can sometimes provide information on exposure risk to different fishing gears and help identify species at high bycatch risk in a fishing area.

Interviews (Surveys conducted with fishermen)

Surveys conducted with fishermen are a relatively inexpensive means of collecting information on bycatch events. These surveys can serve as a first approach to gain an impression of the scale of bycatch impact on PETS and fisheries involved in a specific region before decisions are taken to implement more detailed but expensive monitoring methodologies as direct observations. Limitations of interviews are that they are based on fishermen´s memory or interpretation of events. In the case of some PETS their skill in species identification also needs to be considered. The public are sensitive to the issue of bycatch of many PETS and there is thus an incentive to misreport bycatch rates by fishermen, which needs to be taken into account.

There is a considerable experience of using these survey methodologies in recreational fisheries and small scale fleet data collection programmes with quite high success. Quality issues are addressed in these programmes and experience gained could be very useful for bycatch data collection. In general, these surveys depend on fishers' recall and willingness to volunteer valid information. Recall bias is related to the difficulty of the fisher to remember past events. Non response can also be an issue and should be recorded. In general, like in all self-reporting methods, it is recommended that the information collected is validated through some other on-site sampling methods, such as observers, CCTV cameras, etc.

RCGs together with experts in survey designs and monitoring should be the responsible for dealing with the coordination and harmonisation of the implementation of these statistically sound surveys.

Fishery logbooks

This is a census based methodology. Official logbooks are used to report official data or transversal data under the control regulation (Council Regulation (EC) No. 1224/2009) by the EU fleet (for vessels above 10m LOA). Detailed information is collected on the catch and effort. The introduction of a mandatory file (e.g. as in some Member States) under this regulation to collect PETS data could be considered. However, as in the case of interviews with fishermen, the reporting of data relies on the cooperative spirit and awareness of the fishermen. There are examples where fishery logbooks have been shown to be inconsistent with data collected by independent observations such as onboard scientific observers onboard.

Discard and research survey programmes

Monitoring of bycatch in discard or fishery research programmes can be close to dedicated or direct bycatch monitoring. In this report such sampling is included among indirect methodologies due to the fact that in many cases these programmes are not focused on the collection of bycatch data. Personnel onboard under these programmes have other priorities that could impact on their ability to carry out effective bycatch monitoring. However, these survey personnel can be trained in identification of bycatch species and reporting of fishery data may be expected to be of high quality and can provide an opportunity to extrapolate observed bycatch events to the entire fishery or fleet. The inclusion of a checkbox in the sampling protocols, to report whether the haul was actually checked for incidental bycatch is an important issue to be considered in this type of sampling as it allows the identification of hauls/sets with zero bycatch.

Summary of indirect methodologies

A large quantity of data on bycatch can be collected using indirect methodologies. Logbooks can provide detailed data on catch and effort data with good precision and low coefficient of variation (CV). However, precise estimates may be misleading in terms of accuracy of bycatch estimates if they are based on biased assumptions. There are also concerns about the reliability and representativeness of bycatch data obtained from indirect methodologies. When indirect methods are used to provide bycatch estimates these concerns should be taken into account. Ideally more than one method should be applied and a comparison of the results may help to evaluate and optimise the best practice for monitoring and estimating bycatch in a particular fishery.

As mentioned previously, some of these indirect observation survey methods have been used for the data collection of recreational and small scale fisheries. The expertise gained in this field also addressed quality issues which could be adopted for bycatch data collection programmes

2.4.2 Direct observation methodologies

Direct observations provide very good quality data, but are more expensive than indirect methods. These methodologies need to be used on directed bycatch monitoring programmes where data needs are considered as essential by end-users.

At-sea onboard observer programmes

Independent observations made by trained scientific observers are the most reliable and useful means of collecting data. Many, at-sea, monitoring programmes are undertaken around the world where scientific observers are onboard vessels operating in different fisheries. The objectives of these programmes are diverse as are the tasks covered. A specific, at-sea, scientific observer programme focused on collecting bycatch data would be the ideal way to collect the data required. However these programmes can be expensive, especially for incidental by catch species which only appear rarely in a sampled haul or trip.

CCTV or Electronic Monitoring (EM)

Electronic monitoring (EM) technology consist of multiple closed circuit television cameras, a variety of sensors including Global Position System (GPS), winch rotation and hydraulic pressure, all connected to a video and data storage box. The EM is designed to operate autonomously and continuously while a fishing vessel is at-sea. EM based fisheries monitoring has been carried out over the past decade in a wide range of geographical locations, fishing gears and fishing vessels.

In terms of protected species, EM can be more cost effective, as significant coverage levels can be required for detection of these rare events. Large conspicuous protected species are most easily detected while it can be difficult to detect bycatch events such as seabirds in trawl catch.

EM programmes are more complex than observer programmes. Essential elements of an EM programme include: equipment supply, responsive field services for keeping EM programmes operational on the fleet, data interpretation services for production of standard fishery data from sensor and image data sets, services for consolidating result from EM with other data sources, and an overall management structure to coordinate among all elements of the programme. It is important to note that all these elements need to be taken into account when calculating the cost of electronic monitoring programmes.

Reference fleet

The reference fleet could be a small group of fishing vessels that provides detailed information about their fishing activities. The sampling and data collection methods are similar to the system used by scientific observers onboard on, at-sea, sampling programmes. In the reference fleet the sampling is carried out by a trained crew. The training for the crew is very important to get reliable data, as well as the building of a trust based co-operation between fishermen and scientist. However, as in any self-sampling sampling scheme, it is important to validate the collected data by comparing them with alternative sources of information. Examples of a successful sampling scheme based on a reference fleet can be found in Norway.

In the case of collecting bycatch data, high bycatch risk fisheries fleet vessels should be selected based on gears, fishing activity and geography. The selection of these vessels can be done regionally and coordinated by the RCGs trying to obtain the best coverage in the fisheries selected.

Self-sampling

Self-sampling programmes are developed to get fishermen to self-sample their catches. Using fishermen to collect data on bycatch could be cost effective. Currently there is ongoing effort worldwide to develop programmes to use fishermen to self-sample their catches (ICES 2008b). These programmes may complement the data obtained from other bycatch monitoring programmes. Self-sampling programmes permit an increase in the coverage of the selected fisheries in a cost effective way.

When fishermen are used to collect bycatch data, selection of station and protocols for biological sampling need to be conducted according to proven statistical principles. This should be done training fishermen selected with this aim.

An important issue in self-sampling programmes is the need for incentives for fishermen to participate. If there are no incentives it is more than probable that motivation will be lost and cooperation will decrease over time.

Summary of direct observation methodologies

Bycatch data collected under direct methodologies could provide high resolution and good quality data if based on statistically sound sampling designed monitoring programmes. However, these methodologies are considered as high cost methodologies, and their selection needs to be based on analysis of their cost effectiveness.

2.4.3 Selection of methodology

The choice of the method used for collecting bycatch data in a particular fishery is based on several factors, including:

- Completeness: do the data cover entire range (temporal, spatial, fleet segments etc.) of the fisheries that interact with the species of concern?
- Cost: Is the method cost-effective?
- Timeliness: How quickly are the data available to fisheries scientist, managers, fishermen and other end-users?

- Safety: how safe is the data collection method compared with other monitoring methods and what safeguards are in place to ensure the safety of the data collectors?
- Logistics: How easily is the programme implemented and maintained?
- Planned use of data: do the management goals require a level of detail, quality and timeliness that only certain data sources can provide?

The decision on the selection of the methodologies to used, needs to be analysed at regional scale and case by case by the RCGs taking into account the end-users needs and the factors mentioned above.

2.4.4 Analysis of different data collection methodologies: strengths and weakness

In the previous sections different methodologies used to collect bycatch data have been mentioned and described. An analysis of the strengths and weaknesses of each methodology should be carried out (Table 11). This analysis will help in the selection of the method depending of the objective of the sampling programme.

Table 11. PETS bycatch monitoring methods (strength and weakness).

Methodology	Strengths	Weakness
At-sea scientific observers	1. Detailed bycatch information can be gathered 2. Biological information can be undertaken at-sea 3. Strengthens collaboration among scientist and fishermen 4. Enables data collection on fishing practices too	1. High cost 2. Observer effect (change in fishermen behaviour) is possible. 3. Personal safety risk for observers 4. Limitation to put observers on Small scale vessels due to lack of space or safety reasons 5. In many cases impossible to monitor whole hauls. Subsamples are observed.
REM or CCTV	1. Can record data and imagery from all hauls 2. Reviewers could analyse whole hauls and not subsamples 3. 100% coverage of the whole fleet, hauls is possible. 4. Good alternative in the case of small scale fisheries monitoring 5. Data are available for subsequent quality assurance review 6. High speed data transfer	1. High/Medium cost 2. Expertise needed to analyse images 3. Big amount of data to analyse and storage 4. Not possible to collect biological data 5. Some limitation in the installation of the cameras due to logistic issues 6. Difficulties in the identification of some rare species (mainly fishes due to their small size, how the fish is sorted in the belt etc.)

	7. Opportunity to involve and motivate fishermen in the process	
Self-sampling	1. Low cost 2. All vessels could be monitored 3. Opportunity to involve and motivate fishermen in the process	1. Data quality and reliability not always the best 2. Data collected needs to be uploaded to the data bases by scientific personal which increases the human resources needed 3. Generates additional task for fishermen 4. Incentives to fishermen
Reference fleet	1. Detailed information 2. Data collected similar to scientific observers data (trained crew, using protocols etc.)	1. Low coverage 2. Incentives to fishermen
Interviews to fishermen	1. Low cost 2. High coverage	1. Data quality and level of resolution low 2. Underreporting is a known issue 3. Training for interviewers is needed
logbook	1. Low cost 2. High coverage 3. High speed data transfer	1. Data quality and reliability not always the best 2. Underreporting is a known issue

Taking into account the strengths and weaknesses of the different methodologies, for fisheries where the information is scarce, indirect and low cost methods could be the best options as an initial approach to assessing their impact on PETS.

In the case of well-known and high risk fisheries, a combination of direct methods could be the best option, such as a combination of scientific observers and CCTV or REM.

All these decisions as it has been mentioned in previous sections should be decided under the umbrella of the RCGs, analysing regional specific situations case by case.

2.5 Data availability, data gaps and fisheries characterization

2.5.1 Data availability

Determining an appropriate bycatch sampling programme is clearly dependent on data availability. Bbycatch data is quite scarce, not standardised and not included in a specific data base that could be used for the design of regional sampling programmes, as it is the case for example for commercial fisheries catches and biological data. The Regional Data Base (RDB) is the data base used with this aim by several RCGs (North Sea & Eastern Atlantic, North Atlantic and Baltic).

At present, probably the main data source available on bycatch in the ICES area of protected species is the WGBYC database which holds the data from the yearly reports on the implementation of Council Regulation (EC) No. 812/2004. When information regarding PETS bycatch events has been collected from the DCF, at-sea, monitoring programmes it is included in this data base. However, this information is not available for the RCGs yet and it is also quite unbalanced with respect to cetacean species bycatch data. Probably other ICES working groups where bycatch on PETS is an important issue (i.e. JWGBIRD) could have their own data bases with bycatch information.

Many MS and research institutes have also dealt with bycatch monitoring under different pilot studies and specific bycatch surveys where data on PETS bycatch have been collected. This information is probably held in the databases of National research institutes making access to this information difficult. A similar experience with stomach content data (Pinnegar 2014) demonstrates how a lot of information has been collected historically, but it is dispersed and not widely available. Now a specific data base DAPSTOM is trying to record all this stomach information and deposit it, in the ICES DATRAS data base. A lot of effort is needed to harmonise these data, standardise formats etc. and upload all this data into a regional data base (see section 2.9) but the benefits obtained for a regional sampling programmes is relevant. This could be done in stages coordinated by the RCGs.

2.5.2 General data gaps

A number of data deficiencies need to be addressed before a comprehensive regional sampling design for PETS is possible. At species level, total abundance and bycatch rates knowledge between PETS species is important (ICES 2016). For some species or groups of species, quite detailed information (i.e. cetaceans, sea birds abundance) exist but is totally missing from others (i.e. deep water sharks). This information is essential for the discussion on the sampling design, sampling intensity, and data collection method etc.

There are big differences at regional level as regions compared to others are richer in this information because they have undertaken more specific monitoring and pilot studies etc. To solve data gaps problems and start providing solutions, intersessional work is essential between end-users such as ICES WG experts (Working Groups on the Biology and Assessment of Deep-sea Fisheries Resources, Working Group on Elasmobranch Fishes, WGBYC etc.) and different RCGs, to analyse case by case, different needs at regional level. The role of the RCGs as bridge between end-users and data providers is fundamental.

Information collected from the control regulation (Council Regulation (EC) No. 1224/2009) named as transversal data (logbooks, sales notes, Vessel Monitoring System etc.) is the main source of information used in sampling design. At this level, there are important differences between large scale fisheries and small scale and recreational fisheries. The information obtained from large scale fisheries can be considered as complete. All large scale vessels submit logbook, sales notes and VMS data. However, in the case of small scale and recreational fisheries, this information is scarce and also low quality data, and in some cases (commercial vessels under 10m LOA and recreational fisheries) the transversal information is often missing. This means that for small scale and recreational fisheries, other approaches are needed to collect this basic information (see specific report on small scale and recreational fisheries WP3 Deliverable 3.3).

This is especially relevant, because the impact of these fisheries in inshore waters and for some PETS species is significant. In fact, for some coastal PETS species (i.e. Harbour porpoise, some sea birds) the highest bycatches occur when there are interactions between these species and fisheries using gears such as gillnet and longlines (ICES 2016).

Lack of information concerning bycatch issues in these fisheries is something that needs to be tackled bycatch. Possible methodologies and approaches should be considered to obtain the information required, using statistically sound methods to standardise and harmonise data collection at regional level under the umbrella of the RCGs. ICES WGCATCH and WGRFS experts in collaboration with RCG members should deal with these issues. The design and implementation of a regional scheme for data collection of these specific fisheries is extensively covered in the small scale and recreational fisheries report (WP3 Deliverable 3.3) of fishPi project.

Finally in a highest resolution or specific data level needs, soak time data for passive gear has been requested. Usually effort data from transversal data, logbooks, is submitted as days at-sea. In the case of passive gear this effort unit is not very useful because there is not always a relationship between days at-sea and gear soaks time. This is one reason why onboard observers can help. As longs as mesh size and target species are recorded among transversal fleet data, once some sampling has been achieved, it is possible to infer mean soak time for fleet segments based on the observed hauls, because there are patterns in gear use that are métier specific.

This problem was thoughtfully reviewed in the first workshop on transversal variables held in Zagreb in 2015 (JRC 2015) and in the second workshop on transversal variables held in Cyprus in 2016 (JRC 2016).

A change in the Control Regulation is recommended. Soak time could be considered as mandatory effort unit for fisheries, using passive gear and with high risk bycatch ratio.

Bycatch impacts are gear specific. Gear type records collected in logbooks, particularly with respect to nets is not always sufficiently detailed. It is common to skippers not to differentiate gillnets and trammel nets and mesh size in some trips. The impact on PETS species by these two gear types is very different. When reporting passive gear information in logbooks, the correct information on the gear used is essential.

2.5.3 Fisheries characterization

An important consideration is the extent to which the total bycatch of PETS species of interest is attributable to different types of fishery, such as large scale commercial fisheries, small scale and recreational fisheries. Obviously the relative importance of these components can vary considerably depending on the species concerned. Different end users also have differing and sometimes conflicting priorities. The amount and quality of data available from these three fishery components varies widely and strongly impacts how sampling can be designed. In all cases a detailed knowledge of the different fisheries, and sampling methods appropriate for the different countries, needs to be built up at a regional scale to enable appropriate regional sampling designs to be developed.

For bycatch data collection monitoring, high resolution characterization of the different fisheries is an essential step. Under the DCF, fisheries data collection is reported at level 6 (main gear + target assemblage of species + technical measures e.g. mesh size). This level of description is not enough in the case of by-catch sampling programmes. A real and clear example could be the pelagic pair trawlers case. It is well known and documented (ICES 2013b) that this fleet has an important bycatch rate in dolphins in some fishing grounds.

This métier defined at level six could be pelagic pair trawlers (PTM) targeting small pelagic fishes (SPF) or demersal fishes (DEF) and then the corresponding mesh size. However, data collected by scientific observers demonstrate that there are important differences when this métier is targeting specific species. When the target species is sea bass (*Dicentrachus labrax*) the bycatch rate is higher compared to when mackerel (*Scomber scombrus*) or other species are targeted.

This implies that a higher resolution characterization to that obtained at DCF level 6, of the pelagic pair trawler fleet is required to identify the target fishery population, seasonality etc. . This will also improve knowledge of the specific fleet and make sampling design better and more cost/effective.
Some similar examples occur with other métiers, for example gillnets that are defined as the same métier at level 6, but different fleet segments have quite different impacts in the bycatch rate of some PETS species.

Highest risk métiers have been identified at level 4 by fishing ground by WGBYC. A better resolution description of these métiers could be agreed by the main end-users such as WGBYC, experts groups in sampling design such as WGCATCH and the RCGs. This will facilitate better and more cost effective sampling design.

A similar approach was used in the RCM NA in 2010 to describe métiers identified with the same code at level 6. It was understood that more resolved descriptions of métiers was required because although they had the same code, the dynamics, seasonality and discards rates where quite different according to expert knowledge on these métiers. With this aim, some specific templates were prepared by experts and provided to RCM NA members, which described these metiers in detail by MS and fishing ground.

An adapted template for high risk bycatch métiers could be developed by experts (**Error! Reference source not found.**), using similar criteria. This task could be undertaken by RCG members harmonise to improve sampling programmes for PETS bycatch species.

Once these templates were completed and agreed by the different RCGs, the main métiers by fishing ground with high risk of PETS bycatch or group of species could be identified. A risk-based analysis could be conducted (Section 3.1), to identify which fleets to sample, taking into account end-user needs and priorities based on different criteria such as species vulnerability by fishing ground, cost/benefits of the sampling programme needed etc.

MS involved by métier and fishing ground could beidentified and discuss coordination needs in the RCGs meetings. All this work could be done intersessionally between main end-users, experts in sampling design and members from the RCGs.

Métier description template	
RCM	RCM were this metier should be analysed
Fishing ground	RCM agreed fishing grounds
Name of métier:	Level 6
Flag country:	
Date of update:	Last review of the metier characterization
Description of the métier	

<table>
<tr><td>Spatial distribution of the fishing activity of the métier</td><td colspan="3">Highest resolution spatial distribution (add map when possible)</td></tr>
<tr><td>Seasonal pattern of the fishing activity of the métier</td><td colspan="3">Highest resolution</td></tr>
<tr><td>Number of vessels involved in metier by LOA group:</td><td colspan="3">Based on DCF fleet segmentation</td></tr>
<tr><td>Detailed gear types and selectivity and mitigations devices used in métier</td><td colspan="3">Highest detailed gears description (mesh size, number of nets/hooks etc.)</td></tr>
<tr><td>Main target and bycatch species for the métier</td><td colspan="3"></td></tr>
<tr><td rowspan="5">Indicate level of bycatch risk
(Text. e.g. Significant, Insignificant, Occasional high)</td><td>Species</td><td colspan="2">Bycatch risk</td></tr>
<tr><td></td><td colspan="2"></td></tr>
<tr><td></td><td colspan="2"></td></tr>
<tr><td></td><td colspan="2"></td></tr>
<tr><td></td><td colspan="2"></td></tr>
<tr><td></td><td colspan="3"></td></tr>
<tr><td colspan="4">Sampling of the metier</td></tr>
<tr><td>Indicate if this Métier is merged with other metiers for sampling</td><td colspan="3"></td></tr>
<tr><td>Justification for merging:</td><td colspan="3"></td></tr>
<tr><td rowspan="3">Sampling scheme</td><td>Type of sampling</td><td>Sampling frame and primary sampling unit for data collection</td><td>Data collected (retained catch; discarded catch; unsorted catch) and sampling method (concurrent, other)</td></tr>
<tr><td>Observers at-sea</td><td>Yes, by trip</td><td>Retained and discards</td></tr>
<tr><td>REM</td><td></td><td></td></tr>
<tr><td>Indicate if the Métier is associated with particular sampling problems:</td><td colspan="3"></td></tr>
<tr><td>Protocols adapted to monitor PETS or rare species</td><td colspan="3"></td></tr>
</table>

Figure 2. Metier description template.

2.6 Sampling design

The overarching objective of any sampling programme must be to obtain representative samples with an appropriate level of precision.

The data collected from fisheries have a primary function of supporting stock assessments and informing fleet-based management decisions. To this end, the overall aim for a design-based sampling strategy is to:

1. Collect data in a way that accuracy (bias and precision) can be reliably assessed at national and regional level.
2. Ensure that sampling intensity is allocated in a way that would maximize precision at the level where it matters most in the context of assessment of stocks and fisheries.

The use of statistically robust, design-based sampling schemes that are fully documented is a prerequisite for transparency in the data collection-assessment-advice process. Only with the use of such schemes can bias be properly controlled for and the variance associated with different estimates correctly calculated. It also enables the usage of the data in a wider scientific/management community since the applicability of such data is readily apparent.

The important steps for designing a catch sampling survey are the same for all types of surveys of fisheries, including PETS bycatch data collection programmes:

1) Carefully and clearly define the **objectives** of the proposed survey.
2) Determine what **data** are needed to achieve these objectives.
3) Establish if it is possible to collect the appropriate data, and identify feasible **sampling designs and estimators** (including associated precision estimators) that can be employed, and identify the most problematic aspects.
4) Calculate approximately the **sampling intensity** required to obtain the required survey precision, and the resources required.

All aspects of the survey design, objectives and planned data analysis (including data validation and data storage) should be thoroughly documented before the survey is conducted.

These important steps in the case of PETS must be carried out between main end users (i.e. WGBYC, JWGBIRD etc. for step1 and 2) and experts in sampling design (i.e. WGCATCH, WGRFS for steps 3 and 4) under the umbrella of the RCGs making a case by case study taking into account specific regional needs.

Issues related to regional sampling design have been addressed in some detail under Work Package 2 (WP2) of the fishPi project.

2.7 Sampling intensity and optimization

2.7.1 Sampling intensity

The amount of monitoring needed will depend on the regional objectives, estimates and precision levels decided and agreed between main end-users and the RCGs. Resources available will also need to be considered.

Specifically, species-specific reference points or bycatch limits in numbers or biomass are needed for setting up appropriate sampling schemes on regional or other relevant management scale. Temporal resolution is also a key issue. For some species annual estimates may be needed, whereas for others, point estimates or aggregated measures over longer time periods, such as the MSFD cycle may be sufficient.

The size of the monitoring programmes scheme will also depend fundamentally on the size of the fishery or fisheries to sample. It is therefore imperative to understand the fleet structure, its modes of operation and level of fishing effort by fishery (See fisheries description section above).This part of the work will need to be carried out in the RCGs.

2.7.2 Optimization

There are many different data collection programmes in different sea regions around the world. Some of them are specific and focused on PETS bycatch data collection. Examples include the NOAA data collection programmes in the USA and others in Australia etc.

However, this report is based on bycatch data collection sampling programmes in the ICES Atlantic area. An update of different bycatch sampling programmes will be outlined in this section together with possible options for optimisation of analyses.

Optimisation should be undertaken through close coordination between the RCGs, experts on monitoring and end-users, taking into account data needs and cost/benefit implications.

DCF and Regulation 812/2004
In the ICES Atlantic area the main source of data comes from the current DCF regulation and Control Regulation. However, it has been recognised that the current DCF is not optimised to record PETS bycatch data because the at-sea observer programmes are focused on collecting discard data.

Although there is a desire to monitor a broad range of species, covering several taxa, an overarching design that adequately covers all taxa within the DCF is not realistic. When incorporating monitoring of PETS in the new DCF, the emphasis should therefore be on improving onboard sampling protocols to ensure PETS bycatches are captured within the data recording system and to alter downstream data handling systems to ensure bycatch records of PETS are easily accessed by end users. Not just the bycatch records but also the information on the type of sampling employed. It is also important to know if zero observations are really zero or null.

The recognition that it is difficult to design an overall sampling plan that adequately covers all taxa of rare species does not mean that the sampling focus of the new DCF could not be altered to better reflect patterns of PETS bycatch. Sampling under the existing DCF can contribute to the assessment of bycatch of PETS if such species are properly documented but is not sufficient on its own, as currently implemented, to make a full assessment of the impact of fisheries on PETS. For example, in cases where sampling under the DCF is carried out by means of sampling in harbours (market sampling), obviously the bycatch of PETS is not covered at all because PETS bycatches are not routinely returned to shore. Also the coverage of fleets is not tuned to the métier where there is a high risk of PETS bycatch.

One approach to help address some of these issues maybe to use data collected under the DCF or other sources to help identify “hot spots”, such as areas, seasons or métier with relatively high bycatch rates of PETS. Based on initial assessments of the data at this larger scale, relevant Member States or RCG’s may then need to carry out more focussed surveys to fully assess the scale and patterns of PETS bycatch in specific fisheries. This approach would require Member States or RCM’s to identify additional fisheries and/or species requiring sampling.

Council Regulation (EC) No.812/2004 is focused only on fisheries with high risk of bycatch of cetacean species. The observer coverage in this case is focused on some specific fleets and not all

MS are able to conduct these monitoring programmes. This implies that in some cases the quality of the data collected is not adequate or abscent.

Taking into account both regulations, coordination between MS under the umbrella of the RCGs would allow optimisation of resources needed to fulfil end-user PETS bycatch data requirements.

REM bycatch pilot studies

In recent years, Remote Electronic Monitoring (REM) has played an important role as an efficient alternative methodology for PETS bycatch monitoring. Important improvements have been made in this field as shown by different studies carried out in different countries by different institutes. US, Canada and Australia are considered as pioneers in using this methodology but there are studies that have been carried out in Europe.

DTU Aqua Danish institute was the first one to use CCTV cameras to collect fisheries data in Europe. Specific studies carried out on PETS bycatch monitoring, have been used in different fisheries and fleet segments with quite good results.

DTU aqua and IMARES research institutes used REM methodologies to analyse the impact of gillneters on Harbour Porpoise. Thunen research institute in Germany analysed the bycatch in marine mammals and sea birds of longliners and gillneters. AZTI institute in the Basque Country is using CCTV cameras in purse seiners targeting tropical tunas to collect fishery bycatch data. This means that a significant amount of PETS data has already been collected using this methodology and there is considerable experience within the research community of using REM systems.

A combination of scientific observers at-sea programmes and REM methodologies will probably the best approach when directed PETS bycatch monitoring is needed as cost effective and optimised sampling programmes.

2.8 Data collection

2.8.1 Data collection protocols

In any data collection sampling programme it is essential to have documented and agreed sampling protocols. They are also an important tool in communicating to end-users the details of data collection and any bias that may exist in the final data. Historically sampling is organized at National level and used sampling protocols are developed on a national level. Each country has its specific features that result from culture and tradition. These differences among countries results in different sampling protocols by country. Protocols used in bycatch monitoring programmes are no exception.

Harmonisation of best practice sampling protocols should be an important objective carried out under the RCGs with input from the main end-users. Adoption of these procedures by different MS and institutes involved in bycatch monitoring, will result in more standardised bycatch monitoring programmes at regional level.

In the case of PETS bycatch fish species (i.e. shads, rays, some sharks etc.) at-sea observer DCF programmes protocols harmonised at regional level could be used. A compilation of these different protocols used in different ICES regions was explored in Study Group on Practical Implementation of Discard Sampling Plans (ICES 2012a) with the objective of standardise them.

However, current DCF-related onboard sampling programmes mainly target the fleets and fisheries responsible for the largest catches of the main commercial stocks, with frequent emphasis on trawls. These fleet components are not necessarily those responsible for the largest bycatch of PETS. The sampling protocols may also not be optimal for PETS, for example if there is no specific requirement to record PETS such as mammals, birds or turtles that may fall from gear as they are being brought onboard, or are discarded before the catch reaches the observer's work area, or where there is a very low probability of recording very rare fish species due to sampling of small fractions of each catch. Furthermore, national onboard sampling schemes often sample a very small fraction of the total number of trips of a fleet, further reducing the likelihood of observing rare events such as catches of PETS and leading to many zero observations. For example, it is important that the sampling protocols contain a checkbox indicating whether the haul was actually checked for incidental bycatches and in the case of a scan during hauling, an indicator of the percentage coverage. This checkbox would allow the separation of fishing operations with zero bycatch from those where PET bycatch was neither monitored nor noted.

To address this issue, a questionnaire on sampling practices and logging of PETS information into data-bases was developed by WGBYC, WGCATCH and fishPi WP3 members during the WGCATCH 2015 meeting (Annex section).

This questionnaire will provide a starting point for documentation of sampling practices for PETS bycatch and rare fish species. Such documentation will provide an annual check-point on whether MS have implemented some of the best practices for PETS sampling previously proposed and should provide a reference that allows the tracking of sampling methodologies applied at MS-level and their evolution through time. The appropriate expert groups and the RCGs could then analyse this information based on a regional approach.

2.8.2 Practical implementation of combined discard and bycatch sampling

An important stage and sometimes the most difficult part of any sampling programmes is the implementation phase. Once the statistically sound design, methodologies, protocols etc. are agreed, the next stage is the implementation of the sampling plan.

In the case of PETS bycatch monitoring programmes this part is even more difficult in many cases. Some PETS species such as cetaceans are very sensitive species and evoke an important emotional response in the general public. The main consequence is that some skippers are reluctant to take onboard scientific observers arguing, lack of space, safety reason etc. This limits the random selection of some vessels. Sometimes is also common not to allow alternative monitoring methods such as CCTV. This means that in some cases indirect methodologies should be used where it is documented that the quality of the information is lower compared with direct methodologies.

Refusal rates, reason etc. should be documented in the implementation phase to understand possible bias in the information collected. Nevertheless, a representative coverage of the selected fishery is an important objective.

It is important to maintain or establish good relations with the industry and ensure that the collection of these sensitive data does not compromise further relations. Different incentives to encourage the participation of fishermen can also be proposed and may facilitate the implementation of these programmes. Examples exist in different regions and may be applicable elsewhere (i.e. reference fleet in Norway, extra quota in many MS etc.).

Trained and experience observer teams are essential in all at-sea sampling programmes but specially in the case of PETS bycatch species.

2.9 Data management, analysis and quality evaluation

Data management includes the archiving of data and the processes of quality assurance and quality control. Data stored in common formats will ideally constitute the core of bycatch data collection programmes.

2.9.1 Data base

The database is the link between data sources and data users, allowing feedback to ensure that data types, quantity, quality and origin are consistent with end users requirements. Secondly, the data base facilitates the identification and correction of errors generated during the collection of data, as well as the evaluation of the extent to which processed data are accurate, complete and give a true indication of the value of the factors under consideration. Thirdly, the database ensures data consistency and comparability over time, and facilitates the release of timely information.

The main source of bycatch data is collected under the current DCF and Reg.812/2004. ICES WGBYC has been responsible for compiling and storing all the information collected in the database. At the beginning of this process this database was a compilation of Excel files, until a more complex Access data base was created. However, WGBYC continues to cooperate with the ICES Data Centre to make advances towards developing a more comprehensive WGBYC database design and improving access to ICES Data Centre holdings (i.e. Regional Database (RDB) and Inter-catch).

For this purpose there has been a breakthrough in the work of unifying formats, codes to be used (i.e. adding PETS species international codes) between the expert group and ICES Data Centre experts. The RDB is also the tool that some RCMs (NS&EA, NA and Baltic), are using with the objective of coordinating future regional data collection programmes. The SC RDB has made invested significant effort with the same objective, to standardise formats according to statistically sound samplings programmes. Common and standardised format issues have been covered in the fishPi project under WP2.

Thus, the most effective way of working should be a straightforward coordination between WGBYC and SC RDB with ICES acting as a bridge to take advantage of the progress that has already been made.

This process will require some new developments to incorporate PETS data into the RDB for regional sampling coordination. However, for many of the methods available to sample PETS (such as observers, CCTV cameras or self-sampling), the underlying structure of the sampling is not different from an onboard sampling scheme already included in the RDB, where information is taken on the vessel, trip, hauls, catches by species by haul, and disposal of the catches either as discards or as landings. This may facilitate the incorporation of PETS data into the RDB, and provided the data are collected by probability-based sampling, and the probabilities are known, the analysis methods will be relatively straightforward using the types of routines available in the R-survey package. As the RDB is considered an essential tool for future coordination, its funding is a priority.

2.9.2 Data analysis and quality evaluation

Data quality evaluation of fisheries data encompasses the statistical soundness of the sampling design, the outcomes of implementing the scheme, how the data are managed, and how the data are analyzed. In a practical approach, quality evaluation can be distilled down to two core elements:

a) An evaluation of whether the sampling program is designed and implemented, and the data managed and processed, in a way that follows agreed sets of standards. A programme meeting these standards is in principle capable of providing the desired standard for data quality.
b) An evaluation of the quality of the data that have been collected, using diagnostics and quality indicators that identify potential (or known) bias, and those that provide estimates or indexes of achieved precision.

In recent years many workshops have been carried out under the ICES umbrella (Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment, Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment and Workshop on Practical Implementation of statistical sound Catch Sampling programmes) covering aspects related to data analysis and quality evaluation. Best practices guidelines, protocols and quality check standards have been provided. Under the fishPi project a specific work package (WP4) had been focused on these issues. The he output from WP4 is, applicable to PETS bycatch monitoring programmes and it is recommended that this section of the fishPi project report is consulted.

3 Case studies

In this section, two case studies are analysed. The main objective is to try to cover different PETS species. The first case study is based on PETS species as marine mammals and seabirds and the second on bycatches of PETS fish species such as small sharks, and rays etc.

Under the first case study, fishing areas and fisheries where bycatches of PETS species are considered to be causing a conservation threat will be identified, and document the sampling coverage of these fisheries under the different MS observer programmes. A selection of métiers and areas will be used as an example for the practical implementation of a coordinated sampling for bycatches of PETS.

Case study 2 will use real data from logbooks and a national observer programs aiming to identify a general set of steps to set up a regional program for sampling bycatch of protected, endangered and rare species (PETS). It differs from CS1 in regards of the type of PETS targeted; this study focuses on small to medium sized fish and elasmobranch species that may be encountered in the catch fraction commonly handled by at-sea observers.

3.1 Stages in designing a sampling scheme for PETS such as marine mammals and seabirds

3.1.1 Introduction

Although there is a wish to monitor a broad range of species, covering several taxa, an overarching design that adequately covers all is not realistic. One approach to help address some of these issues maybe to use data collected under the DCF or other sources to help identify “hot spots”, such as fishing grounds, PETS species and métiers with relatively high bycatch rates. Based on initial assessments of the data at this larger scale, relevant Member States or RCG’s may then need to carry out more focussed surveys to fully assess the scale and patterns of PETS bycatch in specific fisheries.

This case study follows this approach focusing in the North Atlantic and North Sea areas. High PETS bycatch risk metiers and fishing grounds are identified in these regions, considering different PETS species or groups of species. Then current observers coverage is analysed on these metiers based on DCF at-sea observers National programmes. Finally possible sampling allocation between MS involved in these fisheries is analysed thinking in a possible regional coordinated sampling programme.

It has to be noted that the presented tables and analysis have been performed at the metier level, as this was the information available from WKBYC and DCF reports. Métier are defined as a group of fishing operations targeting a similar (assemblage of) species, using similar gear, during the same period of the year and/or within the same area and which are characterised by a similar exploitation pattern. However, by definition, metiers cannot generally be sampled with known probability since a list of PSUs (Primary Sampling Unit) is not available in advance. Therefore they cannot be used as strata when designing a sampling design (ICES 2011a). They will be used as domains, parts of the total sampled population, for which estimations will be required. This has to be taken into account when interpreting the presented results and designing a regional bycatch sampling plan.

3.1.2 Material and methods

WKBYC developed a methodology to estimate the bycatch risk of different groups of species, based on the métier, fishing effort and abundance in each different fishing region. The group combined this risk with the DCF sampling effort, to provide an index of which areas and fishing gears are most in need of sampling (ICES 2013b). We have taken advantage of the work done in the WKBYC but focusing in the North Atlantic and North Sea regions. Data from Belgium were not included, as effort reported in DCF National Plans 2011-2013 were presented in units that were not comparable. Data from Germany, which could not be used in the WKBYC report, were included in the present work. All tables have been updated accordingly.

The methodology is summarized in Figure 3. Trying to avoid complexity in the methodology, and focus on key tables, we only show tables 7, 8, 9, 10 and 11 as results in the body of the report. However, all tables needed in the process can be found in Annex to this section. Basically, the methodology followed consists on the following steps:

1. The first step is to define the risk of bycatch for each species by each métier. Within the limited time available, WKBYC used expert opinion and a system of three categories (1: low risk, 2: some risk, 3: high risk) (Table 1). Risk here is taken to represent just the likelihood of bycatch and does not signify the population level risk. The following groups of species were defined: lampreys, round fish, turtles, diving birds, surface birds, seals, dolphins, harbour purpose and large whales.
2. The second step is to identify the presence of the species within the different fishing grounds (presence=1; absence = 0; Table 2). The fishing grounds agreed in the corresponding Regional Coordination Meetings were used (Anon. 2009a, Anon. 2009b): AZ=Azores; BB=Bay of Biscay; CS=Celtic Sea; EA=Eastern Arctic; EB=Eastern Baltic; FI=Faroe Islands; IB=Iberian Sea; IS=Irish Sea; MA=Mid-Atlantic; ME=Mediterranean; NS=North Sea and Eastern Channel; SK=Skagerrak and Kattegat; WB =Western Baltic; WC=Western Channel; WI=Western Ireland; WS=Western Scotland.

3. The combination of the species presence matrix (Table 2) and the risk of bycatch for species by each métier (Table 1) results in a potential risk matrix (e.g. Tables 3.1 – 3.13) and indicates which species have a potential risk in which fishing ground. In the WKBYC report only the tables for the Western Baltic, North Sea and eastern Channel, and the Mediterranean were published. We extended these calculations to all fishing grounds in the North Atlantic and North Sea regions (Annex section).
4. Because fishing intensity of the different métiers differs in each region, the fishing effort of the different métiers has to be taken into account. Therefore the forth step is to combine the potential risk matrix (Table 3) with the fishing effort of the different métiers (in days at-sea) by the different fishing grounds (Table 4). To calculate these tables, the effort by métier and area reported in the DCF National Plans 2011-2013 was used (table III.C.1 of National Plans). Reported effort was indexed with five levels of effort from low to high (Table 5). The resulting matrix gives risk index for each species based on the métier, fishing effort and abundance in each different fishing region (Tables 6.1-6.13).
5. Table 7 then presents those index numbers summed across all species for each fishing region and métier. This provides an index of which areas and fishing gears are most in risk of having significant caught as bycatch, and therefore are most in need of sampling.
6. Table 8 presents index numbers summed across all métiers for each fishing region and species. This provides an index of which species in which areas are most in risk of being caught as bycatch.
7. In order to check the relative distribution of monitoring effort in the DCF against the risk by the métier, the risk index by métier at different regions (Table 7) is then combined with the planned effort in the DCF National programs (Table 9). In order to do this the numbers by métier in both tables are expressed as percentages of the total in each area. The differences between these percentages are given in a sampling index shown in Table 10. Positive numbers indicate relative undersampling; negative numbers indicate relative oversampling. Table 10 therefore provides an initial blueprint for determining which métiers in which fishing areas require relatively more monitoring in order to improve estimates or understanding of bycatch across all protected species groups. Table 10 provides a valuable overall view of the needs of sampling, but it is not intuitive and it should be carefully interpreted. It must be remembered that the indexes values are relative. The sum of all sampling index in the same column (fishing ground) equals zero, and their magnitude will depend on (1) the difference between how the risk of bycatch (table 7) and how the sampling (Table 9) is distributed among the different métiers; and (2) the amount of métiers present in the fishing ground. Comparisons in the magnitude of this sampling index between different fishing grounds should be avoided.
8. Finally, for the selected fishing grounds Bay of Biscay and North Sea (BB and NS), a summary table is presented (Table 11) with the risk index (table 7), a categorization of this index (being 0= no risk, 1=low risk, 2=medium risk, 3=high risk, 4=very high risk); and the number of DCF sampled trips (table 9). This table provides the raw numbers in which table 10 is based. Using filters, it may be used to identify métiers with high risk and zero sampling, and to evaluate the sampling effort of the different métiers.

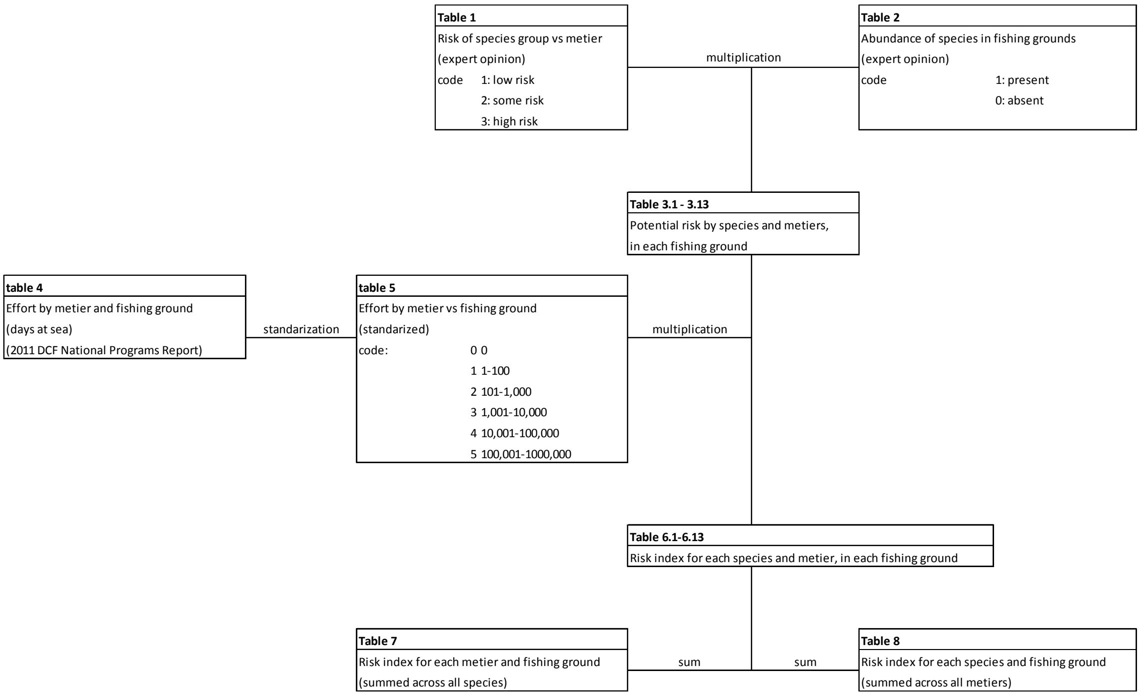


Figure3. Risk based assessment methodology.

Clearly the tables produced in WKBYC, and reproduced here for all fishing grounds in the North Atlantic and North Sea regions, are very dependent on all the assumptions that have been made during the process described above. It is also subject to the improvisations that WKBYC made in determining, among other things, risk of bycatch and population abundance. Fishing effort calculation should also be interpreted with care, as different interpretations may have been used to calculate days at-sea reported in national Programs (JRC 2015). Nevertheless, we agree with WKBYC that the approach, combining species abundance, bycatch rates (or risk), fishing effort and current monitoring levels, is a useful tool to identify the bycatch risk, sampling needs and shortfalls in monitoring. Any further development of this methodology, elaborated in future RCGs, should be taken into account when designing a regional sampling plan.

Regarding species or species groups merging, this is done considering gear effect on the bycatch rate, because of these species specific behaviour in different circumstances (i.e. feeding behaviour of sea surface is very different compared to diving sea birds). Harbour porpoise is considered as species due to their conservation status and impact specific impact of some gears on this species.

3.1.3 Results

Following the methodology developed in WKBYC, a comprehensive set of tables has been produced for the North Atlantic and North Sea regions (Annex section). The aim of these tables is to:

- Provide an overall index of which areas and fishing gears are most in risk of having significant caught as bycatch.
- Provide overall index of which species in which areas are most in risk of being caught as bycatch.
- Evaluate the risk of bycatch of different groups of species, by fishing ground and métiers.
- Evaluate the relative distribution of monitoring effort in the DCF against the risk by the métier for the different fishing grounds, and identify the métiers and fishing grounds which may require more monitoring in order to improve estimates.

Table 3 provides an overall index of which areas and fishing gears are most at risk of having a significant bycatch. This index combines information on the métier, fishing effort and abundance of bycatch species in each fishing ground. It can be observed that some fishing grounds have higher risk of bycatch, than for example the Bay of Biscay (BB), the Iberian Waters (IB), the North Sea (NS) and the Western Channel (WC). There are also significant differences among métiers. Trammel nets and set gillnets present the highest risk factors in all fishing grounds where they are present, followed by otter trawls and set longliners.

Table 3. (Table 7 in methodology section and Annex). Summed risk factors for each métier at different areas (summed across all species). The colour scale from white to dark green is proportional to the risk factor.

	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS	Sum
Boat dredge [DRB]	0	27	18	0	0	36	27	0	32	16	36	18	27	**237**
Bottom otter trawl [OTB]	0	60	60	36	30	75	60	45	60	48	60	60	60	**654**
Multi-rig otter trawl [OTT]	0	52	52	0	13	13	13	0	40	30	26	39	39	**317**
Bottom pair trawl [PTB]	0	39	13	20	13	52	0	0	30	20	13	13	26	**239**
Beam trawl [TBB]	0	24	36	0	0	48	36	0	36	18	36	12	12	**258**
Midwater otter trawl [OTM]	0	51	34	45	17	17	34	34	45	30	51	51	51	**460**
Pelagic pair trawl [PTM]	0	51	34	15	0	34	34	17	45	30	51	51	34	**396**
Hand and Pole lines [LHP] [LHM]	32	27	27	0	0	36	9	0	24	24	27	18	9	**233**
Trolling lines [LTL]	0	36	0	0	0	0	0	0	0	0	24	0	0	**60**
Drifting longlines [LLD]	0	30	15	0	0	45	0	0	0	0	15	30	0	**135**
Set longlines [LLS]	42	60	45	12	0	60	15	0	36	24	45	60	45	**444**
Pots and Traps [FPO]	0	44	44	0	0	44	44	22	50	40	44	33	44	**409**
Fykenets [FYK]	0	45	0	0	0	0	0	0	14	42	0	0	0	**101**
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Trammelnet [GTR]	0	84	63	0	0	105	21	0	72	54	63	42	21	**525**
Set gillnet [GNS]	54	84	63	36	21	105	63	0	72	54	84	63	63	**762**
Driftnet [GND]	0	75	25	0	0	0	0	0	66	22	50	0	0	**238**
Purse-seine [PS]	27	30	10	18	0	40	10	0	18	9	20	10	10	**202**
Lampara nets [LA]	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Fly shooting seine [SSC]	0	0	22	0	0	0	22	0	30	10	22	22	22	**150**
Anchored seine [SDN]	0	0	0	0	0	0	11	0	30	30	11	0	0	**82**
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	**0**
Beach and boat seine [SB] [SV]	0	0	11	0	0	0	0	0	0	20	11	0	0	**42**
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	**0**
Sum	**155**	**819**	**572**	**182**	**94**	**710**	**399**	**118**	**700**	**521**	**689**	**522**	**463**	**0**

Table 4 shows an overall index of which species in which areas are most in risk of being caught as bycatch. As in Table 3, this index combines information on the métier, fishing effort and abundance of bycatch species in each fishing ground. Overall risk for each species or group of species, summed across all fishing ground, ranges from 561 (harbour porpoise) to 809 (round fish). However, some differences can be found in the species at highest risk, depending on the fishing ground. Turtles and diving birds present the highest risk in the Azores. Round fish and turtles have been identified as the species with highest risk in the Bay of Biscay, Celtic Sea, Faroe Islands, Iberian waters, Irish Sea, Mid-Atlantic, Western Channel, Western Ireland and Western Scotland. For the North Sea and Eastern Channel as well as the Skagerrak and Kattegat, round fish and seals are the species with the highest risk.

Table 4. (Table 8 in methodology section and Annex). Summed risk factors for each species at different areas (summed across all métiers). The colour scale from white to dark green is proportional to the risk factor.

	LAMPREYS	ROUND FISH	TURTLES	DIVING BIRDS	SURFACE BIRDS	SEALS	DOLPHINS	HARBOUR PORPOISE	LARGE WHALES	Sum
AZ	13	19	25	22	19	19	19	0	19	**155**
BB	78	103	121	95	89	93	78	77	85	**819**
CS	59	74	90	62	58	62	53	54	60	**572**
EA	19	31	0	23	23	25	22	18	21	**182**
FI	10	14	17	9	9	11	8	8	8	**94**
IB	68	86	115	80	72	78	67	70	74	**710**
IS	44	53	59	42	39	45	38	37	42	**399**
MA	13	17	17	11	14	14	11	8	13	**118**
NS	86	110	0	89	80	92	79	79	85	**700**
SK	70	83	0	66	57	71	56	57	61	**521**
WC	66	89	100	78	74	76	67	66	73	**689**
WI	50	68	82	58	58	56	48	46	56	**522**
WS	49	62	72	49	48	50	43	41	49	**463**
Sum	**625**	**809**	**698**	**684**	**640**	**692**	**589**	**561**	**646**	

It should be highlighted that these indexes are very dependent on all the assumptions made during the process described in the methodology of this Case Study. They are therefore subject to the improvisations that the WKBYC made in determining the risk of bycatch and population abundance by expert opinion. It would be advisable that values based on expert opinion are double checked by experts in each fishing ground. In particular, we found the values for turtles to be surprisingly high for some fishing grounds as the Bay of Biscay.

Once we estimate the risk of each métier by fishing ground, it is interesting to compare it with the distribution of sampling effort onboard.Table 5 provides a sampling index which may help in the identification of which métiers are relatively under or over sampled. The index is relative and should be interpreted with care, but the table clearly shows a relative over-sampling of bottom otter trawlers in all fishing grounds. This can be easily explained because this fleet is usually covered by onboard sampling programmes designed to sample discards. As the DCF covers the sampling of discards but not of bycatch of PETS, the same effect is observed in other métiers (i.e. trammel netters in the Bay of Biscay, or set gillneters in the Celtic Sea).

Table 5. (Table 10 in methodology section and Annex). Relative DCF sampling effort (Table 9 in methodology section and Annex) subtracted from relative summed risk factors (Table 7 in methodology section and Annex) for each métier at different areas. Positive numbers (in green), indicate relative under sampling; negative numbers (in red) indicate relative over sampling.

	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	0	3	0	0	0	5	1	0	2	3	2	3	6
Bottom otter trawl [OTB]	0	-28	-36	-24	-50	-47	-32	-12	-39	-63	-29	-21	-57
Multi-rig otter trawl [OTT]	0	6	9	0	14	2	3	0	4	6	4	7	8
Bottom pair trawl [PTB]	0	-2	2	-4	14	-5	0	0	4	4	2	2	6
Beam trawl [TBB]	0	3	-10	0	0	4	8	0	-13	3	-16	2	3
Midwater otter trawl [OTM]	0	-4	6	-14	18	2	2	6	-1	6	4	-40	2

Pelagic pair trawl [PTM]	0	3	2	6	0	4	6	14	6	6	5	2	3
Hand and Pole lines [LHP] [LHM]	-29	3	5	0	0	5	2	0	3	5	4	3	2
Trolling lines [LTL]	0	4	0	0	0	0	0	0	0	0	3	0	0
Drifting longlines [LLD]	0	4	3	0	0	6	0	0	0	0	2	6	0
Set longlines [LLS]	-23	-2	8	7	0	5	4	0	5	5	7	11	10
Pots and Traps [FPO]	0	5	3	0	0	6	-25	19	6	2	6	3	-7
Fykenets [FYK]	0	5	0	0	0	0	0	0	2	3	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	-21	9	0	0	11	5	0	6	10	-5	8	5
Set gillnet [GNS]	35	10	-12	20	4	4	14	-27	-2	0	-5	7	13
Driftnet [GND]	0	9	4	0	0	0	0	0	9	4	7	0	0
Purse-seine [PS]	17	4	2	10	0	-3	3	0	3	2	3	2	2
Lampara nets [LA]	0	0	0	0	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	3	0	0	0	5	0	0	2	3	3	5
Anchored seine [SDN]	0	-3	0	0	0	0	3	0	3	0	2	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	2	0	0	0	0	0	0	4	2	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

In general, the distribution of DCF sampling effort among the different métiers varies greatly among fishing grounds, both in terms of intensity (numbers of trips sampled) and coverage of the different fleets (Table 6). A regional coordinated sampling scheme for bycatch species should ensure that the métiers with higher bycatch risk are sufficiently covered in each fishing ground. RCGs could prioritize métiers and species or group of species to be sampled taking into account end-users needs, species conservation status etc. RCGs coordinated with experts groups in monitoring such as WGCATCH and WGRFS, can decide methodologies, sampling design etc. In case of bycatch scientific observers at-sea, observers programmes combined with REM methods seems to be the best approach. Case by case cost/benefit analysis will be needed to select the most appropriate regional sampling programme.

Table 6. (Table 9 in methodology section and in Annex). Number of trips sampled on-board by métier for each fishing ground under the DCF as reported by MS National Plans 2011-2013.

MÉTIER	no of trips	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	70			7				31		20		12		
Bottom otter trawl [OTB]	4,149		134	113	17	32	652	266	13	421	425	127	97	145
Multi-rig otter trawl [OTT]	12									12				
Bottom pair trawl [PTB]	183		24		6		141							
Beam trawl [TBB]	316			40			36	6		161		72	1	
Midwater otter trawl [OTM]	524		40		15		4	36	6	62		12	148	18
Pelagic pair trawl [PTM]	125		13	10	1		8	16		6		8	24	9
Hand and Pole lines [LHP] [LHM]	180	144	0	0	0	0	0	0	0	0	0	0	0	0
Trolling lines [LTL]	0													
Driftinglonglines [LLD]	268													
Set longlines [LLS]	252	144	36				36			0				
Pots and Traps [FPO]	384			11			0	202		12	36	0	11	34

Fykenets [FYK]	33									0	30			
Stationary uncovered poundnets [FPN]	102													
Trammelnet [GTR]	432		120	6			40			40		46		
Set gillnet [GNS]	711			56		7	126	8	7	105	60	56	16	2
Driftnet [GND]	0													
Purse-seine [PS]	270						96							
Lampara nets [LA]	15													
Fly shooting seine [SSC]	44			2				3		36			3	
Anchored seine [SDN]	60		12							12	36			
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach-seine [SB]	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

In order to work in the practical design and implementation of a coordinated sampling for bycatches of PETS, two fishing grounds were selected because they presented a high overall summed risk of bycatch (Table 3, Table 4): the Bay of Biscay and the North Sea and Eastern Channel. For each fishing ground, a group of métiers were chosen, together with some species presenting high bycatch risk. This selection also considers their impact in the different PETS groups, trying to cover all of them. Lampreys and round fish groups are not considered in this case study because these PETS species are better covered in Case Study 2.

1. Bay of Biscay fishing ground

Bottom pair trawl and Pelagic pair trawl fleet were selected because of their bycatch of cetaceans (dolphins, harbour porpoise and large whales). Trammel nets and Set gillnets were selected because of their bycatch of cetaceans and diving birds (Table 7).

Table 7. (Table 6.2 in methodology section and in Annex)**.** Identified risk factor for the species groups for each métier in the Bay of Biscay. The colour scale from white to dark green is proportional to the risk factor. Numbers in red show the selected case studies

MÉTIER	BB								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	3	3	3	3	3	3	3	3	3
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	8	8	12	4	4	4	4	4	4
Bottom pair trawl [PTB]	6	6	9	3	3	3	3	3	3
Beam trawl [TBB]	4	2	6	2	2	2	2	2	2
Midwater otter trawl [OTM]	3	9	6	6	6	6	6	3	6
Pelagic pair trawl [PTM]	3	9	6	6	6	6	6	3	6
Hand and Pole lines [LHP] [LHM]	3	3	3	3	3	3	3	3	3
Trolling lines [LTL]	3	3	3	6	9	3	3	3	3
Drifting longlines [LLD]	2	2	6	4	6	2	2	2	4
Set longlines [LLS]	4	4	12	8	12	4	4	4	8
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	9	6	3	6	3	9	3	3	3
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	4	12	12	12	4	12	8	12	8
Set gillnet [GNS]	4	12	12	12	4	12	8	12	8

Driftnet [GND]	3	9	9	9	9	9	9	9	9
Purse-seine [PS]	3	3	3	3	3	3	6	3	3
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

In Table 8 the total effort (days at-sea) can be observed by the selected métiers in the case of the Bay of Biscay. France and Spain are the main contributors to the total effort in this fishing ground. With respect to pair bottom trawlers, Spain is nearly the total effort of this métier and similar to France in the case of trammel nets and pelagic pair trawls. Gillnet effort is distributed mainly between the two MS (France and Spain) with France accounting for 77% of the effort, and United Kingdom with less than 2%.

For métiers where close to the total effort is executed by just one MS there is no need for any coordination. All the responsibility should be allocated to that MS. However, as is the case for gillnets, the sampling design and sampling effort allocation, should be made considering the effort attributed to different MS and their total contribution to the total effort.77% of the total effort executed by France and 20% by Spain in this case. Because of their small contribution to effort, the United Kingdom will not need to sample this fleet.

Table 8. Total effort (days at-sea) in percentage in the Bay of Biscay by métier.

Fleet	ESP	FRA	IRL	NLD	UK	Total general
GNS	20,5%	77,7%	0,0%	0,0%	1,7%	100%
GTR	0,3%	99,7%	0,0%	0,0%	0,0%	100%
PTB	96,3%	3,7%	0,0%	0,0%	0,0%	100%
PTM	0,0%	99,5%	0,0%	0,4%	0,0%	100%

It is interesting to compare the fishing effort performed by the different métiers with their sampling coverage under the current DCF. To explore this, a summary table gathering the information about the summed risk factor by métier and the number of trips sampled onboard, was designed (Table 9). This kind of table provides an easy way to identify the métiers under high risk of by-catch in the Bay of Biscay, and to check their current sampling coverage. For example, by filtering métiers with risk category 3 and 4, which may be interpreted as high bycatch risk, it is observed that there are three métiers which are not sampled by any Member State (multi-rig otter trawls, set gillnets and driftnets). Three métiers presented sampling below 50 trips a year (midwater otter trawls, pelagic pair trawls and set longlines). And, the two remaining métiers, bottom otter trawls and tramelnets, presented a sampling effort above 100 trips a year. This information is in line with the sampling index provided in Table 5 and in our view, both tables are complementary.Table 5provides an overall index of relative under or oversampling, and Table 9goes into the detail for a particular fishing grounds, and shows the absolute numbers.

Focusing on the métiers selected for the Bay of Biscay, it can be observed that the sampling coverage differs greatly depending on the métiers. Trammel nets are covered with 120 trips, bottom pair trawls with 24 and pelagic pair trawls with 13. Set gillneters are not sampled at all. With the bycatch data available it is not possible to recommend a sampling effort (in number of trips) for each of these métiers. This will be a future task under the remit of the RCG and bycatch sampling dedicated working groups.

Table 9. Summary table for the Bay of Biscay. SUM RISK: the summed risk factor for each métier (table 3); RISK CAT: the summed risk factor categorized (0:0, 1:1-25, 2:26-50, 3: 51-75, 4:76-150); DCF SAMP: the number of trips sampled onboard under the DCF (table 6). Numbers in red show the selected case studies

	BB		
	SUM RISK	RISK CAT	DCF SAMP
Boat dredge [DRB]	27	2	0
Bottom otter trawl [OTB]	60	3	134
Multi-rig otter trawl [OTT]	52	3	0
Bottom pair trawl [PTB]	39	2	24
Beam trawl [TBB]	24	1	0
Midwater otter trawl [OTM]	51	3	40
Pelagic pair trawl [PTM]	51	3	13
Hand and Pole lines [LHP] [LHM]	27	2	0
Trolling lines [LTL]	36	2	0
Drifting longlines [LLD]	30	2	0
Set longlines [LLS]	60	3	36
Pots and Traps [FPO]	44	2	0
Fykenets [FYK]	45	2	0
Stationary uncovered poundnets [FPN]	0	0	0
Trammelnet [GTR]	84	4	120
Set gillnet [GNS]	84	4	0
Driftnet [GND]	75	3	0
Purse-seine [PS]	30	2	0
Lampara nets [LA]	0	0	0
Fly shooting seine [SSC]	0	0	0
Anchored seine [SDN]	0	0	12
Pair seine [SPR]	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0
Glass eel fishing	NA	NA	NA

An important precursor to sampling design and sampling effort allocation to MS, is to have a better knowledge of the fleet (see 2.5.3 section). Information on the seasonality of the fleet, fleet segments involved etc. is needed. It is important not just focus on the total effort made by the fleet and each of MS.

Once this is done, a coordinated sampling design can be made taking into account, end-users needs in the species or group of species where it is known that the impact of this fleet is relevant (Cetaceans and diving sea birds in this case). This coordination responsibility is under the umbrella of the RCGs, with close collaboration with the relevant experts (fisheries monitoring experts and bycatch experts in case of PETS).

Another important issue to be considered will be the possibility of the implementation of the agreed coordinated sampling programme. General guidelines are provided in section 2.8 of this report.

2. North Sea & Eastern Channel

Set longlines and hand and pole lines were selected in the North Sea because of their bycatch of surface birds (Table 10)

Table 10 (Table 6.9 in methodology section and Annex). Identified risk factor for the species groups for each métier in the North Sea and Eastern Channel. The colour scale from white to dark green is proportional to the risk factor. Numbers in red show the selected case studies**.**

MÉTIER	NS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	4	4	0	4	4	4	4	4	4
Bottom otter trawl [OTB]	10	10	0	5	10	10	5	5	5
Multi-rig otter trawl [OTT]	8	8	0	4	4	4	4	4	4
Bottom pair trawl [PTB]	6	6	0	3	3	3	3	3	3
Beam trawl [TBB]	8	4	0	4	4	4	4	4	4
Midwater otter trawl [OTM]	3	9	0	6	6	6	6	3	6
Pelagic pair trawl [PTM]	3	9	0	6	6	6	6	3	6
Hand and Pole lines [LHP] [LHM]	3	3	0	3	3	3	3	3	3
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	3	3	0	6	9	3	3	3	6
Pots and Traps [FPO]	10	5	0	5	5	5	5	5	10
Fykenets [FYK]	3	2	0	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	4	12	0	12	4	12	8	12	8
Set gillnet [GNS]	4	12	0	12	4	12	8	12	8
Driftnet [GND]	3	9	0	9	9	9	9	9	9
Purse-seine [PS]	2	2	0	2	2	2	4	2	2
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	6	6	0	3	3	3	3	3	3
Anchored seine [SDN]	6	6	0	3	3	3	3	3	3
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

In Table 11 the total effort (days at-sea) can be observed in percentage, by the selected métiers in the case of North Sea. There are five MS (France, Germany, Netherlands, Sweden and United of Kingdom) involved in the total effort but with important differences considering different métiers. Swedish effort is very small in all métiers so in this case it will not be necessary for this MS to be involved in any monitoring.

It can also be seen that in the case of the three métiers, two MS are the main effort contributors but with important differences in the effort executed. In the case of hand lines (LHM), United Kingdom executes 85% and France with 15% of the effort. In the case of rod and pole lines (LHP), France contributes with 71% and Netherlands with 27% of the effort. Finally, in the case of longlines (LLS), France contributes with 30% and United Kingdom with 67% of the effort.

These métiers have an impact mainly in the surface sea birds. The contribution involves 4 MS and with important differences depending on the métier. In this case co-ordinated sampling effort will be important.

Table 11 Total effort (days at-sea) in percentage in the North Sea by métier.

Fleet	FRA	GER	NLD	SWE	UK	Total general
LHM	14,7%	0,0%	0,0%	0,0%	85,3%	100%
LHP	71,8%	0,1%	27,7%	0,4%	0,0%	100%
LLS	30,5%	0,0%	0,0%	2,0%	67,5%	100%

It is interesting to compare the fishing effort performed by the different métiers and their sampling coverage under the current DCF. As for the Bay of Biscay, a summary table was designed for the North Sea and Eastern Artic, in order to assess the current sampling coverage in this fishing ground. The table gathers information about the summed risk factor by métier and the number of trips sampled onboard (Table 12). This kind of table provides an easy way to identify the métiers with a high risk of by catch in the North Sea, and to check their current sampling coverage. For example, by filtering métiers with risk category 3 and 4, which may be interpreted as high bycatch risk, it is observed that there is one métier which is not sampled by any MS in the North Sea (driftnets). Trammel nets were sampled with a sampling effort of 40 trips, set gillnets with 105 trips, and bottom otter trawls with 421. This information is in line with the sampling index provided in Table 5. As mentioned above, in our viewTable 5 and Table 12 are complementary. The first one provides an overall index of relative under or oversampling, and the second one goes into the detail for a particular fishing grounds, and shows the absolute numbers.

If we focus on the métiers selected for the North Sea and Eastern Artic, it can be observed that the two selected métiers are not sampled by any MS. It should be noted, that although the summed risk for these métiers are not as high as others, they do have a significant impact on certain species as surface birds. With the bycatch data available it is not possible to recommend a sampling effort (in number of trips) for the selected métiers. This will be future task, under the remit of the RCG and bycatch sampling dedicated working groups.

Table 12. Summary table for the North Sea and eastern Artic. SUM RISK: the summed risk factor for each métier (table 3); RISK CAT: the summed risk factor categorized (0:0, 1:1-25, 2:26-50, 3: 51-75, 4:76-150); DCF SAMP: the number of trips sampled onboard under the DCF (table 6). Numbers in red show the selected case studies.

	NS		
	SUM RISK	**RISK CAT**	**DCF SAMP**
Boat dredge [DRB]	32	2	20
Bottom otter trawl [OTB]	60	3	421
Multi-rig otter trawl [OTT]	40	2	12
Bottom pair trawl [PTB]	30	2	0
Beam trawl [TBB]	36	2	161
Midwater otter trawl [OTM]	45	2	62
Pelagic pair trawl [PTM]	45	2	6
Hand and Pole lines [LHP] [LHM]	24	1	0
Trolling lines [LTL]	0	0	0
Drifting longlines [LLD]	0	0	0
Set longlines [LLS]	36	2	0

Pots and Traps [FPO]	50	2	12
Fykenets [FYK]	14	1	0
Stationary uncovered poundnets [FPN]	0	0	0
Trammelnet [GTR]	72	3	40
Set gillnet [GNS]	72	3	105
Driftnet [GND]	66	3	0
Purse-seine [PS]	18	1	0
Lampara nets [LA]	0	0	0
Fly shooting seine [SSC]	30	2	36
Anchored seine [SDN]	30	2	12
Pair seine [SPR]	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0
Glass eel fishing	NA	NA	NA

As in the case of the Bay of Biscay fishing ground, understanding the way the fleet operates is an essential precursor toregional sampling design and effort allocation (see 2.5.3 section). Why is this previous fleet knowledge essential? As an example we can select the LHP métier from the North Sea. It seems quite clear that France is the main contributor in effort and Netherlands is the second MS in importance. This is based on total year effort contribution. It seems evident that the sampling effort allocation should be proportional to the total effort contribution by MS. Now, we go to the scenario where the bycatch risk rate is higher in one season compared with others (i.e. some sea surface birds breeding season). After a higher resolution characterization of this métier, it is identiied that it is in the breeding season that the Netherlands executes all the effort and France is 20% only. In this case both MS effort is quite similar and this season is where more sampling is needed. In this case the effort allocation will be similar for the two MS, when *a priori* it was expected that France should be adopt the main responsibility for monitoring this fleet. This scenario occurs in the real world. Hence, the importance of this type of analysis. This work as it is described in section 2.5.3 could be done by MS experts, trying to characterise their fleets in the RCGs. Following the Bay of Biscay example, regional design, sampling effort allocation, and methodologies etc. could be carried out.

3.1.4 Conclusions

Risk based assessment can be considered as a good approach to help identify “hot spots” as fishing grounds and métiers with high bycatch risk on different PETS species or groups of species. The methodology explained in this case study could be carried out by the RCGs, as a prior step before implementing regional sampling programmes to collect bycatch data on PETS.

As highlighted in the case study, this methodology is based in some assumptions (such as species abundance by fishing ground, or bycatch risk by métier). These assumptions could be progressively removed with the improvement of bycatch data collection, if the information required can be by completed with good quality data.

This methodology identifies MS involved in the different métiers by fishing ground and the monitoring or sampling coverage carried out. In this case only current DCF coverage by observer programmes has been analysed but this can be done with all methodologies and sampling programmes used to collect PETS bycatch data (Council regulation (EC) No.812/2004).

Regional sampling programmes should be designed following the stages explained in section 2. The risk assessment should be carried out under the umbrella of the RCGs with contribution from main end-users to identify data needs and experts in fisheries monitoring for the sampling design. The in-depth description of the fisheries, together with the sampling design, the identification of implementation problems and the development of guidelines for good practice are also an important task to be undertaken regionally, to ensure that collected data can be merged and used at a regional level to fulfil end-user needs. These tasks will need to involve a multidisciplinary team working intersessionally.

3.2 Pilot study in "An approach to evaluate sampling design for monitoring PETS in at-sea observer programmes"

3.2.1 Introduction

In Case Study 2 we used data from the Swedish national at-sea observer program to exemplify an approach to evaluate alternative designs when setting up a national or regional programme for sampling bycatch of protected, endangered and threatened species (PETS). Case Study 2 differs from CS1 in regards of the type of PETS targeted, focusing on small to medium sized fish and elasmobranch species that may be encountered in the catch fraction commonly handled by at-sea observers. The case study draws on the work from Work Package 2 (WP2) but was not based on the trip based data on landed catches compiled for WP2 because occurrence of PET species in commercial catches is frequently low and/or irregular and it therefore seemed worthwhile to explore the uses of at-sea samples from national observer programmes. Most ongoing at-sea sampling programmes under the DCF collect multi-species data and aim to obtain quarterly estimates of discards. However, due to budgetary and other constraints the number of observed trips is usually limited, resulting in over-stratified designs and poor estimates, particularly in uncommon species. This is also the case for the Swedish national observer program. In the present case study observer data were therefore collapsed over several years. A simulated population was then created and resampled to investigate alternative stratification designs and effort allocation for monitoring bycatch of rare species. Spurdog (*Squalus acanthias*) was used as focus species in the case study because it had both reasonably good logbook data and at-sea observer data. A small number of other 'rare species' were also analysed, alongside a subset of sensitive species identified as candidate species for assessment of OSPARS biodiversity indicator FC1: *Population abundance of a suite of selected species*. To keep track on eventual changes in sampling precision in these prioritised species in the simulations we also included three important and frequent commercial species for which biological information is always sampled by the at-sea observers. Species-specific fishery description based on historical logbook data was used to identify alternative stratification designs with regards to season and area.

3.2.2 The Swedish demersal trawl fishery in Skagerrak and Kattegat (ICES Division 3a)

The Swedish demersal trawl fisheries in Kattegat and Skagerrak (Area 3aS and 3aN, respectively) can roughly be divided into four fleets based on mesh size and selection device: the *Pandalus* fishery (32-69 mm mesh size) with or without sorting grid, the directed *Nephrops* fishery (70-89 mm mesh size) with sorting grid, and the mixed demersal fishery (<90 mm mesh size) targeting cod and other demersal species. The *Nephrops* fishery and the mixed demersal fishery operate throughout the area, but the *Pandalus* fishery is more or less restricted to the Skagerrak, the deeper northern part of division 3a. From 2000 – 2014 the four fleets made approximately 182,000 fishing trips in Division 3a (Table 13). For the purpose of this study fishing trips were assigned to quarter and area based on total landed weight using a simple dominance rule.

Table 13. Number of fishing trips by fleet, area and quarter in the Swedish demersal trawl fishery 2000-2014 from logbook records.

	Area	*3an*				*3as*			
Fleet	*Season*	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
OTB_CRU_32-69_0_0_all		5631	7394	5632	5758	[1]			
OTB_CRU_32-69_2_22_all		3584	4996	5421	3654	[1]			
OTB_CRU_70-89_2_35_all		6596	9612	14044	7049	1568	4171	5102	1959
OTB_CRU_90-119_0_0_all		13140	14636	13454	10852	8883	8165	10987	9670
Area × Season		28951	36638	38551	27313	10451	12336	16089	11629

[1] A small number of *Pandalus* trips was reported from the northern Kattegat (3as) but was not included in the simulations.

3.2.3 At-sea observer data

The present Swedish at-sea observer programme is stratified based on the combination of fleet × area × quarter resulting in 24 national strata. Although the sampling programme has been modified over time it has more or less followed the same stratification scheme throughout the time period 2000-2014 with the exception of the *Pandalus* fishery that was not included in the regular sampling programme before 2008. The programme aims to sample 3 fishing trips per strata and year, i.e., 72 fishing trips per year (48 fishing trips per year before 2008). Consequently, under perfect sampling circumstances, 816 fishing trips would be expected to have taken place during whole period 2000-2014. However, logistic issues and variable response and refusal rates reduced the number of fishing trips actually observed in the demersal trawl fisheries to 747 observed fishing trips (Table 14).

Table 14 Number of observed fishing trips by fleet, area and quarter in the sample.

	Area	*3an*				*3as*				
Fleet	*Season*	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	*Fleet total*
OTB_CRU_32-69_0_0_all		11	14	10	11					46
OTB_CRU_32-69_2_22_all		9	10	8	9					36
OTB_CRU_70-89_2_35_all		28	43	32	38	18	31	25	17	232
OTB_CRU_90-119_0_0_all		59	60	43	45	60	50	55	61	433
Area × Season total		107	127	93	103	78	81	80	78	747

3.2.4 Auxiliary data

Landings of spurdog (*Squalus acanthias*) from 1996 to 2010 were used to investigate spatial and temporal variation of spurdog catches in Division 3a. Until 2007 there was no specific legislation regulating the fishery for spurdog in Swedish waters, but from 2008 onwards the fishery was restricted to licensed vessels, and from 2011 onwards landing of spurdog was prohibited. Between 1996 and 2008 more than 95% of the landed spurdog was caught in two fisheries: the mixed demersal trawl fishery (OTB_CRU_90-119_0_0_all) and a shark gillnet fishery (Figure 4). In the mixed demersal trawl fishery most of the landings came from the fourth quarter whereas the directed gillnet fishery landed similar quantities in the second and fourth quarter. Smallest landings were observed during the first quarter in both fisheries.

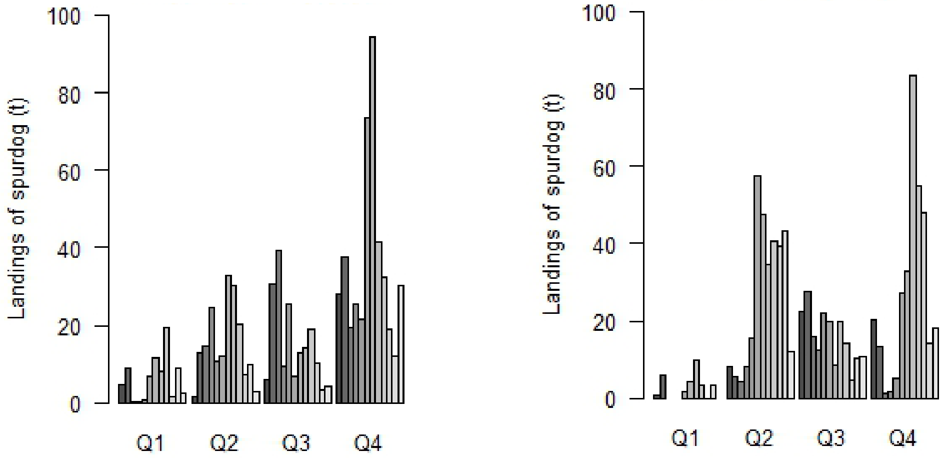


Figure 4. Quarterly landings of spurdog, *Squalus acanthias*, (in tonnes) from 1996 to 2008 in two fisheries contributing more than 95% of the landed weight (Left graph, trawl; right graph, gillnet).

The landings from the mixed demersal trawl fleet (Figure 5) came from a much more widespread area than the landings from the gillnet fishery (Figure 6), but in both fisheries most of the landed fish were caught relatively close to the Swedish Skagerrak coast where 2-3 'hot-spots' could be visually identified.

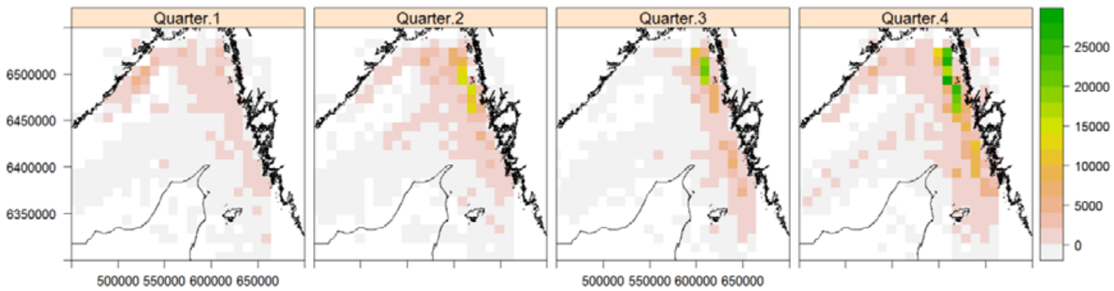


Figure 5. Swedish landings of spurdog (kg) by quarter in the mixed demersal trawl fishery 1996 – 2008, based on logbook data.

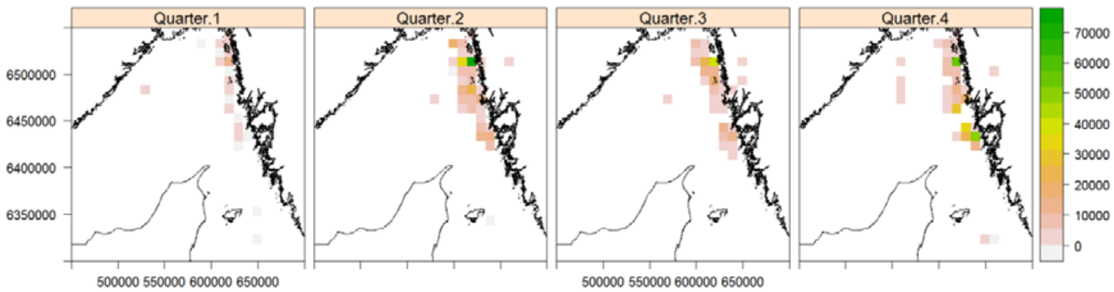


Figure 6. Swedish landings of spurdog (kg) by quarter in the directed shark gillnet fishery 1996 – 2008, based on logbook data.

3.2.5 Simulation and estimation

Trip-level data were used to simulate the different scenarios. Haul level information and length data from at-sea observations were available but were not used in the present simulation. For spurdog, and the rare and sensitive species, weight of the landing and discard fractions in each trip were combined into total catch. For the selected commercial species, only the weight of the landed fraction was used and compared with the official data on landings for these species.

The following steps were taken to build the data set used in simulations:

1) An artificial population of *N* fishing trips was created from the trips observed in the at-sea observer programme. Within each stratum, the observed trips with associated data were replicated until the number of fishing trips for that stratum in the logbook data was reached (*i*.e., the number of observed trips from Table 14).
2) To simulate different stratifications, stratum boundaries were re-arranged into new strata according to the design of interest.
3) Resampling was carried out using simple random sampling without replacement within each stratum.
4) Estimation was carried out in the R-package survey (Lumley, 2014).
5) Step 3 to 5 was repeated 500 times for each design, each of those simulations being used to generate an estimate of the total catch, or, in the case of the commercial species, the landed fraction, using the Horvitz-Thompson estimator.
6) Based on the 500 estimates, mean estimated total catch, standard errors and percentile confidence intervals were calculated.

The following stratification/effort allocation designs were simulated;

1. The fleet × area × quarter stratification presently used in the Swedish at-sea programme. Because the *Pandalus* fleets mainly operate in the Skagerrak, this adds up to 24 strata.
   1.1. Equal effort allocation with sample size n = 3 samples per stratum, in total 72 samples, as in the present national programme.
   1.2. Equal effort allocation with sample size n = 30 samples per stratum, in total 720 samples which approximates the total number of observed trips in the data set.
   1.3. Unequal effort allocation with sample size n = 10 for the *Pandalus* fleets and n=25 and n=55 for the *Nephrops* and mixed demersal fleets, respectively. In total 720 samples.

2. Based on the spatial and temporal distribution of spurdog landings Stratified by fleet and quarter. In the simulation the two *Pandalus* fleets were combined which resulted in 12 strata (the three fleets; *Pandalus*, *Nephrops* and Mixed demersal and 4 quarters).
   2.1. Equal effort allocation with sample size n=6 samples per stratum, in total 72 samples.
   2.2. Unequal effort allocation with sample size n=20 for the *Pandalus* fleets and n=50 and n=110 for the *Nephrops* and mixed demersal fleets, respectively. In total 720 samples.

3.2.6 Results & Discussion

Design 1.1., the sampling design presently used in the Swedish at-sea programme resulted in very high variability in estimated total catch of the rare and sensitive species. The relative standard error for spurdog (1.33) and halibut (1.87) were much higher than for the other species in the two groups (Figure 7). Increasing the sample size to n=30 (design 1.2) reduced the standard error for spurdog and halibut considerably (to 0.41 and 0.56, respectively) and the remaining species showed relatively low variability, ranging from 0.11 to 0.27 (Figure 8). Optimizing effort allocation would potentially reduce estimate variability further, but the effect on the relative standard error (RSE) of the simulated unequal effort allocation (design 1.3) was small (-1.9% for the rare, and -7.5% for the sensitive species on average, not shown).

Reducing the number of strata and increasing the number of samples within the new strata (design 2.1 not shown, and design 2.2) did not improve the precision of the catch estimates; on average the relative standard error (RSE) generated by the new stratification/effort allocation was 1.5 times higher than in the present design (Figure 9).

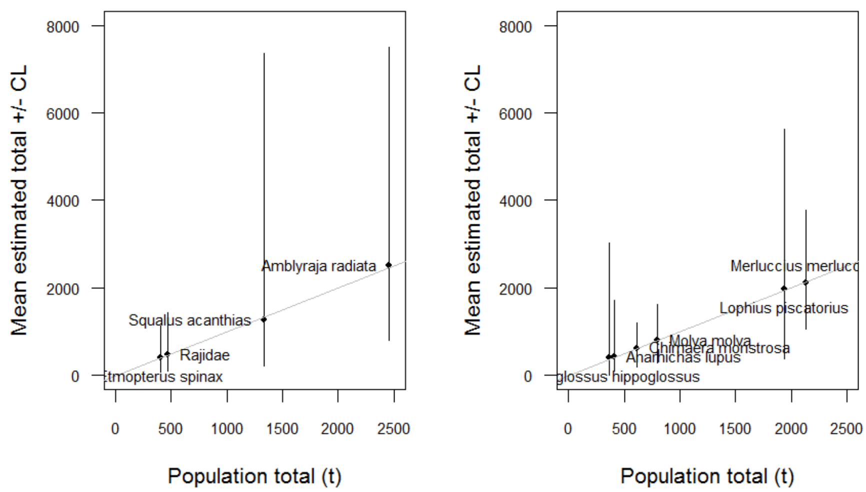


Figure 7. Design 1.1, 24 strata, equal effort n=3, total sample size = 72.

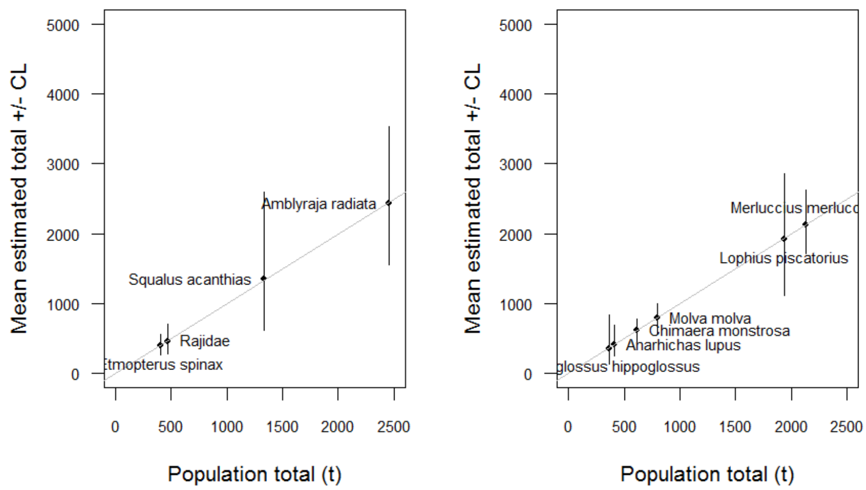


Figure 8. Design 1.2., 24 strata, equal effort n=30, total sample size = 720.

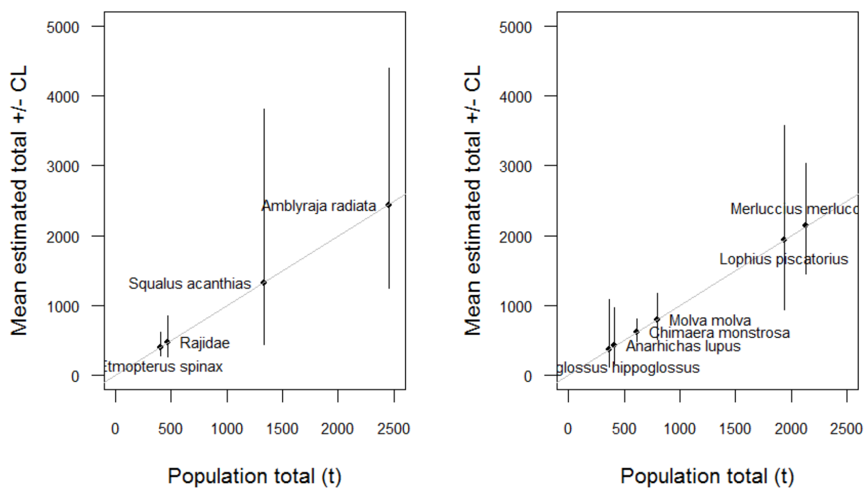


Figure 9. Design 2.2., 12 strata, unequal effort, total sample size = 720**.**

Comparison of estimated total landings of the three commercial species also suggests that the present stratification scheme generates narrower confidence limits than the alternative design (Figure 10). The confidence limit are still very wide however, which is a consequence of the small sample size (Figure 10, top panel). A summary of the simulation result based on the design and effort allocation of the present at-sea observer programme (design 1.1) is presented in Table 15.

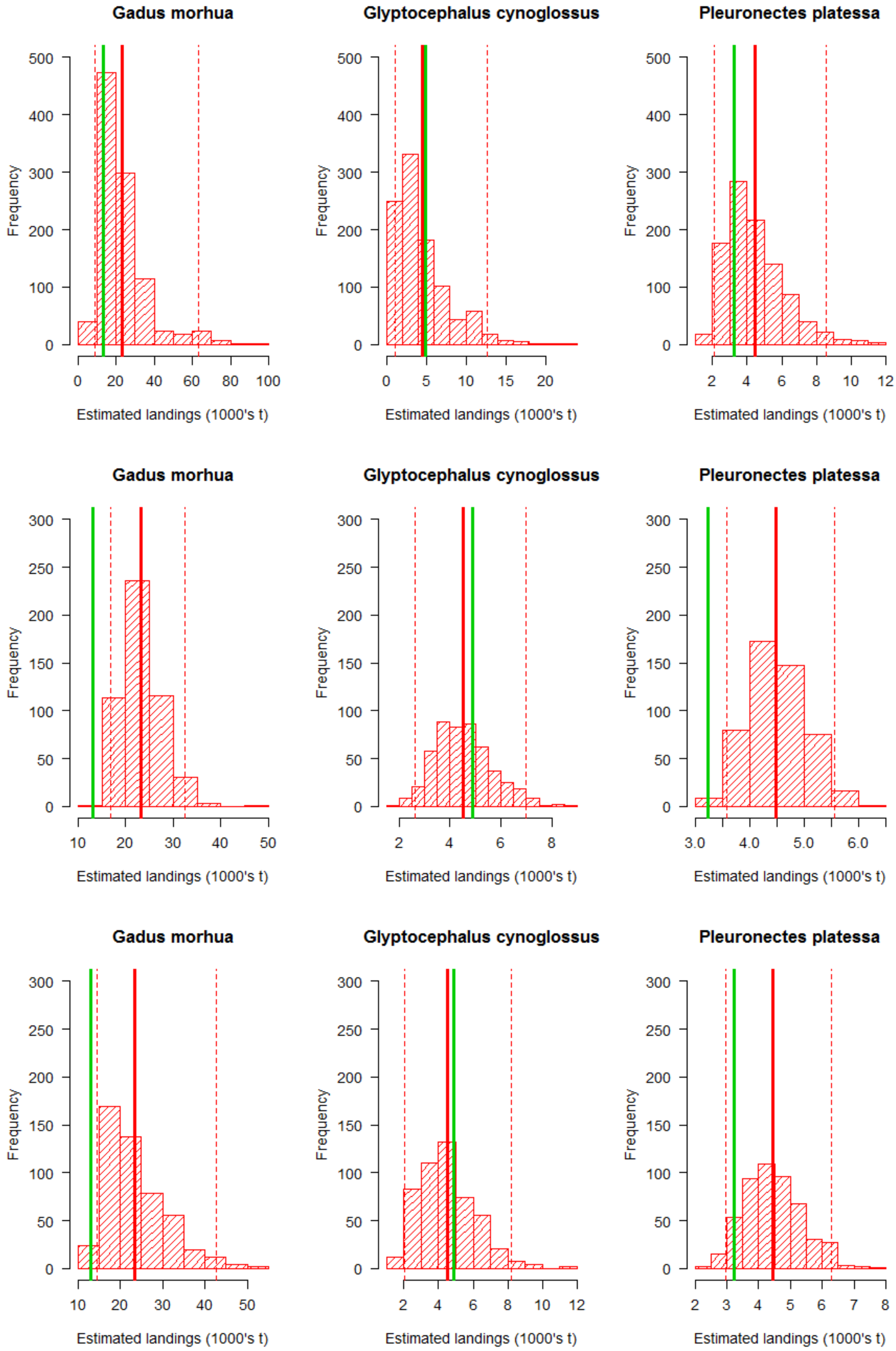


Figure 10 Distribution of estimated landings (1000's tonnes) of three commercial species between 2000 and 2014. Mean, lower and upper 95% confidence limit in red, the total landings from logbook data in green. Top

panel, design 1.1, 24 strata, equal effort n=3, total sample size = 72. Middle panel, design 1.2., 24 strata, equal effort n=30, total sample size = 720. Bottom panel, design 2.2., 12 strata, unequal effort, total sample size = 720.

Table 15. Mean estimated catch or estimated landing in tonnes (t) from 2000-2014 for three groups of elasmobranchs and fish. For spurdog and three commercially important species total landings (t) from logbook are also presented. Mean estimates, relative standard error (RSE) and percentile confidence limits were based on estimates from 500 simulations.

Species group		***Mean estimated***	***Total***	*RSE*	***95% CI***	
Local name	***Scientific name***	***total catch (t)***	***landings (t)***		***Lower***	***Upper***
PETS (Catch)						
Starry ray	*Amblyraja radiata*	2513		0.67	813	7498
Velvet belly	*Etmopterus spinax*	404		0.64	78	1133
Skates, rays[1]	*Rajidae*	476		0.73	110	1449
Spurdog	*Squalus acanthias*	1274	729[1]	1.33	229	7361
Sensitive (Catch)						
Catfish	*Anarhichas lupus*	431		0.91	99	1724
Rabbit fish	*Chimaera monstrosa*	627		0.42	192	1201
Halibut	*Hippoglossus hippoglossus*	402		1.82	20	3022
Anglerfish	*Lophius piscatorius*	1967		0.76	387	5620
Hake	*Merluccius merluccius*	2115		0.34	1065	3773
Ling	*Molva molva*	805		0.45	285	1633
Commercial (Lan)						
Cod	*Gadus morhua*	23068	13104	0.53	8720	63118
Witch	*Glyptocephalus cynoglossus*	4458	4899	0.76	1101	12731
Plaice	*Pleuronectes platessa*	4437	3238	0.38	2115	8590

[1] 643 tonnes of spurdog was landed in the demersal trawl fishery from 2000 – 2007. Between 2008 and 2010 86 tonnes was landed.

3.2.7 Conclusions

National at-sea sampling programmes serve multiple purposes and should not be changed without considerable evaluation. The presented approach can be used to guide decisions on new or expanded monitoring programmes with goals such as optimising sampling of single species or groups of species in mind. For example, if additional information for specific species is need over a management cycle it could be possible to allocate additional effort in some of the present strata to obtain the optimal allocation evaluated in simulations. It would also be possible to create new strata based on similar exercises. As mentioned earlier, most at-sea sampling programmes collect multi-species data, but are limited by small sample sizes. Collapsing data over several years is one way of increasing sample sizes for simulation and preliminary explorative analyses and for some purposes, reliable estimates of total catch over longer time periods may be sufficient. However, for final analyses and estimation it is recommended to include regional data on disaggregated level, similar to the approach in WP2.

4 Conclusions

As a first step to PETS regional coordination data collection programmes, substantial work needs to be carried out between main end users, data collectors and experts in sampling, under the umbrella of the RCGs, to: i) identify priorities on data needs and criteria; ii) identify additional national requirements for data to support national inshore management schemes, and find a trade-off between regional and national needs; iii) identify appropriate data collection methods for each country taking into account cost/benefits analysis and practical implementation considerations; and iv) develop standardised guidelines and protocols. This work will lead into a second phase of development to design a regional data collection programme to cover needs identified in the first phase. This should be fully documented and submitted by the RCG for peer review by experts in survey design. The process outlined implies considerable intersessional work between end-users and data providers, led by RCGs, and for which some funding and resources will be needed. Expertise and training is needed on data collection and analysis methods and sharing of expertise and skills across countries.

Although there is a desire to monitor a broad range of species, covering several taxa, an overarching design that adequately covers all taxa within the DCF is not realistic. When incorporating monitoring of PETS in the new DCF/ DC MAP, the emphasis should therefore be on improving onboard sampling protocols to ensure PETS bycatches are captured within the data recording system and to alter downstream data handling systems to ensure bycatch records of PETS are easily accessed by end users.

One approach to help address some of these issues maybe to use data collected under the DCF or other sources to help identify “hot spots”, such as areas, seasons or métiers with relatively high bycatch rates of PETS. Based on initial assessments of the data at this larger scale, relevant MS or RCG’s may then need to carry out more focussed surveys to fully assess the scale and patterns of PETS bycatch in specific fisheries. This approach would require MS or RCM’s to identify additional fisheries and/or species requiring sampling.

An important consideration is the extent to which the total bycatch of PETS species of interest is attributable to different types of fishery, such as large scale commercial fisheries, small scale and recreational fisheries. Obviously the relative importance of these components can vary considerably depending on the species concerned. Different end users also have differing and sometimes conflicting priorities. The amount and quality of data available from these three fishery components varies widely and strongly impacts how sampling can be designed. In all cases a detailed knowledge of the different fisheries, and sampling methods appropriate for the different countries, needs to be built up at a regional scale to enable appropriate regional sampling designs to be developed.

When selecting the methodology to use, cost benefit analysis and strengths and weaknesses should be also considered taking into account different criteria (e.g. cost, data quality etc.).

A combination of scientific observers at-sea programmes and REM methodologies will be probably the best approach when directed PETS bycatch monitoring is needed as cost effective and optimised sampling programmes.

A common database is an essential tool for regional coordination (i.e. RDB). This should be the preferred option to host PETS data. A lot of work has been invested in recent years to standardise formats and protocols to upload this data into this database in different RCGs, ICES workshops, and SC RDB etc.

Two case studies are provided in the report: 1) risk based assessment and; 2) an approach to evaluate sampling design for monitoring PETS in at-sea observer programmes. Case Study 1, risk based assessment, can be considered as a good approach to help identify “hot spots” as fishing grounds and métiers with high bycatch risk on different PETS species or groups of species. The methodology explained in this case study could be carried out by the RCGs, as a prior step before implementing regional sampling programmes to collect bycatch data on PETS. Case Study 2 can be used to guide decisions on new or expanded monitoring programmes with goals such as optimising sampling of single species or groups of species in mind.

References

ASCOBANS. 2011. Report of the ASCOBANS Bycatch Working Group (BYCWG). AC18/Doc.4-07.

ASCOBANS. 2013. Report of the ASCOBANS Bycatch Working Group (BYCWG). AC20/Doc.3.1.1.a.

ASCOBANS. 2014a. Report of the ASCOBANS Bycatch Working Group (BYCWG). AC21/Doc.3.1.1.a rev.1.

ASCOBANS 2014b. A preliminary report from the Joint ACCOBAMS/ASCOBANS Working Group on the Marine Strategy Framework Directive (MSFD). ASCOBANS AC21_13.3.1

ASCOBANS. 2015. Report of the Workshop on Remote Electronic Monitoring with Regards to Bycatch of Small Cetaceans. ASCOBANS, The Hague, Netherlands, 2 October 2015. WS/REM/2015 Report.

Babcock, E. A., and E. K. Pikitch. 2003. How much observer coverage is enough to adequately estimate bycatch? Pew Institute for Ocean Science and Oceana, 36 p.

Cochran, W. G. (1977), Sampling Techniques. John Wiley & Sons, New York.

Council Regulation (EC) No 812/2004 of 26 April 2004 laying down measures concerning incidental catches of cetaceans in fisheries and amending Regulation (EC) No 88/98.

Council Regulation (EC) 199/2008 of 25 February 2008 concerning the establishment of a Community Framework for the collection, management and use of data in fisheries sector for scientific advice regarding the Common Fisheries Policy.

Commission Regulation (EC) No 665/2008 of 14 July 2008 laying down detailed rules for the application of Council Regulation (EC) No 199/2008 concerning the establishment of a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy.

Commission Regulation (EC) No 1078/2008 of 3 November 2008 laying down detailed rules for the implementation of Council Regulation (EC) No 861/2006 as regards the expenditure incurred by Member States for the collection and management of the basic fisheries data.

Council Regulation (EC) No. 1224/2009 of of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy.

Commission Decision (EC) No 2010/93/EC of 2010 adopting a multi annual Community programme pursuant to Council Regulation (EC) No 199/2008 establishing a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy.

Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.

Council Directive 2008/56 EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive).

ICES. 2008a. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU), 27–30 October 2008, Bergen, Norway. ICES CM 2008\ACOM:32. 41 pp.

ICES. 2008b. Report of the Workshop on Fishers Sampling of Catches (WKSC), 10–13 June 2008, ICES, Copenhagen, Denmark. ICES CM 2008/ACOM:30. 61 pp.

ICES, 2009. Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE), 8-11 September 2009, Copenhagen, Denmark. ICES CM 2009/ACOM:40. 617 pp.

ICES 2010. Report of the ICES Workshop on methods for merging metiers for fishery-based sampling (WKMERGE). ICES CM 2010 / ACOM:40.

ICES 2011a. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 8-10 November 2011, Bilbao, Spain. ICES CM 2011/ACOM:52

ICES. 2011b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2011 /ACOM:50.

ICES. 2011c. Report of the Joint NAMMCO/ICES Workshop on observation schemes for bycatch of mammals and birds (WKOSBOMB, 28 June-1 July 2010, ICES, Denmark. ICES CM 2010/ACOM: 33. 40 pp.

ICES 2012a. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2012 /ACOM:50

ICES 2012b. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS 2), 6-9 November 2012, Copenhagen. ICES CM 2012/ACOM:54

ICES. 2013a. Report of the Working Group on Bycatch of Protected Species (WGBYC), 4–8 February, Copenhagen, Denmark. ICES CM 2013/ACOM:27. 73 pp.

ICES. 2013b. Report of the Workshop on Bycatch of Cetaceans and other Protected Species (WKBYC), 20-22 March 2013, Copenhagen, Denmark. ICES CM 2013/ACOM: 36. 55pp.

ICES. 2013c. Report of the Workshop on Data Collection - assessments of non-fishery impacts (WKDCF-NF), 8 -10 October 2013, ICES Headquarters, Copenhagen, Denmark. ICES CM2013/ACOM:74. 70 pp.

ICES. 2014a. Report of the Working Group on Bycatch of Protected Species (WGBYC), 4–7 February 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:28. 96 pp.

ICES. 2015a. Report of the Working Group on Bycatch of Protected Species (WGBYC), 2-6 February 2015, ICES Headquarters, Copenhagen, Denmark. ICES CM 2015\ACOM:26. 82 pp.

ICES. 2015b. Report of the Working Group on Commercial Catches (WGCATCH), 10-14 November 2014, ICES HQ, Copenhagen, Denmark. ICES CM 2014\ACOM:36. 295 pp.

ICES. 2016. Working Group on Bycatch of Protected Species (WGBYC), 1–5 February 2016, ICES HQ, Copenhagen, Denmark. ICES CM 2016/ACOM:27. 617 pp.

JRC. 2015. Report on the 1st Workshop on Transversal Variables. 19-23 January 2015. Zagreb, Croatia. 82pp

JRC. 2016. Report on the 2nd Workshop on Transversal Variables. 22-26 February 2016. Nicosia, Cyprus. 108pp

Kindt-Larsen, L., Dalskov, J., Stage, B., and Larsen, F., 2012. Observing incidental harbour porpoise *Phocoena phocoena* bycatch by remote electronic sensing. Endangered Species Research, 19:75–83.

McElderry, H. 2008. At-sea observing using video-based electronic monitoring. Archipelago Marine Research Ltd., 55 p.

NMFS. 2004. Evaluating bycatch: a national approach to standardized bycatch monitoring programs. U.S. Dep. Commer., NOAA Technical Memorandum NMFSF/SPO-66, 108 p.

Northridge, S. and L. Thomas, 2003. Monitoring levels required in European Fisheries to assess cetacean bycatch, with particular reference to UK fisheries. Final report to the Department for Environment Food and Rural Affairs, European Wildlife Division. August 2003.

Northridge, S. P. & Fortuna, C. M. Aug 2008 Report of the INTERNATIONAL WORKSHOP ON BYCATCH WITHIN THE ACCOBAMS AREA. General Fisheries Council of the Mediterranean, p. 1-37.

Northridge, S. 2011. An overview of the state of bycatch monitoring and mitigation measures being implemented in European fisheries. IWC SC/63/SM21.

Pinnegar, J.K. 2014. Rebuilding the DAPSTOM database to enhance compatibility with ICES datasets. Final Report for EU Contract No MARE/2012/02 – SI2.632887. Centre for Environment, Fisheries & Aquaculture Science, Lowestoft, UK. November 2014, 29pp.

Scientific, Technical and Economic Committee for Fisheries (STECF) Revision of DCF (STECF-14-02). 2014. Publications Office of the European Union, Luxembourg, EUR 26573 EN, JRC89196, 103 pp.

Scientific, Technical and Economic Committee for Fisheries (STECF) – DCF Revision - part 4 (STECF-14-07). 2014. Publications Office of the European Union, Luxembourg, EUR 26612 EN, JRC 89788, 77 pp.

T. Lumley (2014) "survey: analysis of complex survey samples". R package version 3.30.

Annex section

Annex 1. Questionnaire on PETS protocol

	Indicate: Y/N/NA*	Does the on board sampling protocol ask to record this information?									Is the National database designed to enter this information?								
	Member state	EE	FI	LV	LT	PL	BG	HR	CY	GR	EE	FI	LV	LT	PL	BG	HR	CY	GR
1	Does the protocol contain instruction to record catch of other vertebrate species than fish (i.e. turtles, birds, dolphins, seals)?																		
2	In gill nets - and hook-and-line fisheries: does the protocol instruct to indicate how much of the hauling process has been observed for (large) incidental bycatches which never came on board (because they fall out of the net)?																		
3	Does the protocol contain a check for rare specimens in the catch at opening of the codend or immediate removal during hauling in gill nets or hook-and-line?																		
4	If Yes: is the observer instructed to indicate if the codend was not checked in a haul or at how much of the hauling process has been checked for immediate removal?																		
5	Does the protocol instruct to check for rare specimens during sorting of the catch (i.e. at conveyor belt)?																		

6	If Yes: is the observer instructed to indicate how much of the sorting process has been checked on “haul level” (i.e. percentage)?																		
7	Does the protocol instruct to report specific handling or devices on board which may hide incidental bycatch?**																		
8	If Yes: is the observer instructed to report what effect this has on the sampling at “haul level”?																		
9	Does the protocol instruct to report of mitigation (i.e. Acoustic Deterrent Devices or “pingers”)?																		
10	If yes for ADD’s: is there a check for proper working (i.e. Battery check)?																		
11	In case of an incidental catch: is the observer instructed to indicate its state (dead and discarded, released alive, discarded in unknown state, collected for further research?																		

Annex 2. Tables produced for risk based assessment methodology (cs1)

Table 11. Identified risk (by expert opinion) for species groups by each fishing métier: 1: low risk; 2: medium risk; 3: high risk.

GEAR	LAMPREYS	ROUND FISH	TURTLES	DIVING BIRDS	SURFACE BIRDS	SEALS	DOLPHINS	HARBOUR PORPOISE	LARGE WHALES
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 2. Abundance of the species group (by expert opinion) at the different areas: 1: present; 0: absent

AREA	LAMPREYS	ROUNDFISH	TURTLES	DIVING BIRDS	SURFACE BIRDS	SEALS	DOLPHINS	HARBOUR PORPOISE	LARGE WHALES
AZ	1	1	1	1	1	1	1	0	1
BB	1	1	1	1	1	1	1	1	1
CS	1	1	1	1	1	1	1	1	1
EA	1	1	0	1	1	1	1	1	1
FI	1	1	1	1	1	1	1	1	1
IB	1	1	1	1	1	1	1	1	1
IS	1	1	1	1	1	1	1	1	1
MA	1	1	1	1	1	1	1	1	1
NS	1	1	0	1	1	1	1	1	1
SK	1	1	0	1	1	1	1	1	1
WC	1	1	1	1	1	1	1	1	1
WI	1	1	1	1	1	1	1	1	1
WS	1	1	1	1	1	1	1	1	1

Table 12.1 Potential risk for the species groups for each métier in the Azores

MÉTIER	AZ								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whal es
Boat dredge [DRB]	1	1	1	1	1	1	1	0	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	0	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	0	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	0	1
Beam trawl [TBB]	2	1	3	1	1	1	1	0	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	0	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	0	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	0	1
Trolling lines [LTL]	1	1	1	2	3	1	1	0	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	0	2
Set longlines [LLS]	1	1	3	2	3	1	1	0	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	0	2
Fykenets [FYK]	3	2	1	2	1	3	1	0	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	0	1
Trammelnet [GTR]	1	3	3	3	1	3	2	0	2
Set gillnet [GNS]	1	3	3	3	1	3	2	0	2
Driftnet [GND]	1	3	3	3	3	3	3	0	3
Purse-seine [PS]	1	1	1	1	1	1	2	0	1
Lampara nets [LA]	1	1	1	1	1	1	1	0	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	0	1
Anchored seine [SDN]	2	2	1	1	1	1	1	0	1
Pair seine [SPR]	2	2	1	1	1	1	1	0	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	0	1
Glass eel fishing	2	1	1	1	1	1	1	0	1

Table 3.2. Potential risk for the species groups for each métier in the Bay of Biscay

Métier	BB								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.3. Potential risk for the species groups for each métier in the Celtic Sea

MÉTIER	CS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.4. Potential risk for the species groups for each métier in the Eastern Artic

MÉTIER	EA								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	0	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	0	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	0	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	0	1	1	1	1	1	1
Beam trawl [TBB]	2	1	0	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	0	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	0	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	0	1	1	1	1	1	1
Trolling lines [LTL]	1	1	0	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	0	2	3	1	1	1	2
Set longlines [LLS]	1	1	0	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	0	1	1	1	1	1	2
Fykenets [FYK]	3	2	0	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	0	1	1	2	1	1	1
Trammelnet [GTR]	1	3	0	3	1	3	2	3	2
Set gillnet [GNS]	1	3	0	3	1	3	2	3	2
Driftnet [GND]	1	3	0	3	3	3	3	3	3
Purse-seine [PS]	1	1	0	1	1	1	2	1	1
Lampara nets [LA]	1	1	0	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	0	1	1	1	1	1	1
Anchored seine [SDN]	2	2	0	1	1	1	1	1	1
Pair seine [SPR]	2	2	0	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	0	1	1	1	1	1	1
Glass eel fishing	2	1	0	1	1	1	1	1	1

Table 3.5. Potential risk for the species groups for each métier in the Faroe Island

MÉTIER	FI								
	lampreys	**round fish**	**turtles**	**diving birds**	**surface birds**	**seals**	**dolphins**	**harbour porpoise**	**large whale s**
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.6. Potential risk for the species groups for each métier in the Iberian waters

MÉTIER	IB								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whal es
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.7. Potential risk for the species groups for each métier in the Irish Sea

MÉTIER	IS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.8. Potential risk for the species groups for each métier in the Mid-Atlantic

MÉTIER	MA								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.9. Potential risk for the species groups for each métier in the North Sea and Eastern Channel

MÉTIER	NS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	0	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	0	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	0	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	0	1	1	1	1	1	1
Beam trawl [TBB]	2	1	0	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	0	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	0	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	0	1	1	1	1	1	1
Trolling lines [LTL]	1	1	0	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	0	2	3	1	1	1	2
Set longlines [LLS]	1	1	0	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	0	1	1	1	1	1	2
Fykenets [FYK]	3	2	0	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	0	1	1	2	1	1	1
Trammelnet [GTR]	1	3	0	3	1	3	2	3	2
Set gillnet [GNS]	1	3	0	3	1	3	2	3	2
Driftnet [GND]	1	3	0	3	3	3	3	3	3
Purse-seine [PS]	1	1	0	1	1	1	2	1	1
Lampara nets [LA]	1	1	0	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	0	1	1	1	1	1	1
Anchored seine [SDN]	2	2	0	1	1	1	1	1	1
Pair seine [SPR]	2	2	0	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	0	1	1	1	1	1	1
Glass eel fishing	2	1	0	1	1	1	1	1	1

Table 3.10. Potential risk for the species groups for each métier in the Skagerrak and Kattegat

MÉTIER	SK								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	0	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	0	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	0	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	0	1	1	1	1	1	1
Beam trawl [TBB]	2	1	0	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	0	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	0	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	0	1	1	1	1	1	1
Trolling lines [LTL]	1	1	0	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	0	2	3	1	1	1	2
Set longlines [LLS]	1	1	0	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	0	1	1	1	1	1	2
Fykenets [FYK]	3	2	0	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	0	1	1	2	1	1	1
Trammelnet [GTR]	1	3	0	3	1	3	2	3	2
Set gillnet [GNS]	1	3	0	3	1	3	2	3	2
Driftnet [GND]	1	3	0	3	3	3	3	3	3
Purse-seine [PS]	1	1	0	1	1	1	2	1	1
Lampara nets [LA]	1	1	0	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	0	1	1	1	1	1	1
Anchored seine [SDN]	2	2	0	1	1	1	1	1	1
Pair seine [SPR]	2	2	0	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	0	1	1	1	1	1	1
Glass eel fishing	2	1	0	1	1	1	1	1	1

Table 3.11. Potential risk for the species groups for each métier in the Western Channel

MÉTIER	WC								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.12. Potential risk for the species groups for each métier in the Western Ireland

MÉTIER	WI								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 3.13. Potential risk for the species groups for each métier in the Western Scotland

MÉTIER	WS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	1	1	1	1	1	1	1	1	1
Bottom otter trawl [OTB]	2	2	3	1	2	2	1	1	1
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	1	1	1	2	3	1	1	1	1
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	2	1	1	1	1	1	1	1	2
Fykenets [FYK]	3	2	1	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	1	1	1	1	1	2	1	1	1
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	1	1	1	1	1	1	1	1	1
Fly shooting seine [SSC]	2	2	1	1	1	1	1	1	1
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	2	2	1	1	1	1	1	1	1
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	2	1	1	1	1	1	1	1	1

Table 4. Total effort (days at sea) in the reference fleets by métier (level 4) for each Fishing Ground (all member states combined without Belgium).

MÉTIER	Effort	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	124,530		8,193	735			49,595	5,795		28,981	139	26,439	219	4,436
Bottom otter trawl [OTB]	445,176		40,712	24,693	2,332	763	100,395	20,664	2,458	104,875	37,970	32,288	38,109	39,920
Multi-rig otter trawl [OTT]	59,380		31,461	10,328		3	13	35		10,268	3,370	624	1,082	2,197
Bottom pair trawl [PTB]	22,863		1,962	1	335	2	12,691			7,466	197	37	66	106
Beam trawl [TBB]	112,581		153	5,112			13,028	1,134		86,212	292	6,585	65	2
Midwater otter trawl [OTM]	15,109		2,159	322	1,159	31	21	978	383	5,493	438	1,057	1,615	1,455
Pelagic pair trawl [PTM]	12,946		5,443	138	31		408	122	11	2,787	458	2,076	1,044	430
Hand and Pole lines [LHP] [LHM]	71,134	32,650	4,212	4,626	0	0	16,083	63	0	6,192	1,464	5,560	278	9
Trolling lines [LTL]	2,173		1,593									580		
Driftinglonglines [LLD]	7,858		430	28			7,251					47	102	
Set longlines [LLS]	107,347	7,315	21,659	1,525	18		51,968	68		3,017	209	5,131	10,658	5,781
Pots and Traps [FPO]	375,097		20,203	14,699			89,661	22,566	111	105,203	13,392	44,039	2,862	62,364
Fykenets [FYK]	7,821		1,514							19	6,288			
Stationary uncovered poundnets [FPN]	0													
Trammelnet [GTR]	175,059		32,127	1,274			111,896	32		21,033	1,079	7,489	130	1
Set gillnet [GNS]	277,909	3,076	23,074	7,041	345	91	157,137	1,031		49,602	7,803	20,472	6,643	1,595
Driftnet [GND]	4,123		2,316	43						1,401	12	351		
Purse-seine [PS]	78,126	5,085	3,797	41	124		68,081	6		498	89	397	1	9
Lampara nets [LA]	0													
Fly shooting seine [SSC]	8,169			738				124		6,366	71	305	255	312
Anchored seine [SDN]	4,217							2		1,333	2,879	3		
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach-seine [SB]	339			12							323	5		
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 5. Fishing effort by métier at the different areas by fishing days. (0:0, 1: 1–100, 2: 101–1000, 3: 1001–10 000, 4: 10 001–100 000, 5: >100 000)

MÉTIER	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	0	3	2	0	0	4	3	0	4	2	4	2	3
Bottom otter trawl [OTB]	0	4	4	3	2	5	4	3	5	4	4	4	4
Multi-rig otter trawl [OTT]	0	4	4	0	1	1	1	0	4	3	2	3	3
Bottom pair trawl [PTB]	0	3	1	2	1	4	0	0	3	2	1	1	2
Beam trawl [TBB]	0	2	3	0	0	4	3	0	4	2	3	1	1
Midwater otter trawl [OTM]	0	3	2	3	1	1	2	2	3	2	3	3	3
Pelagic pair trawl [PTM]	0	3	2	1	0	2	2	1	3	2	3	3	2
Hand and Pole lines [LHP] [LHM]	4	3	3	0	0	4	1	0	3	3	3	2	1
Trolling lines [LTL]	0	3	0	0	0	0	0	0	0	0	2	0	0
Driftinglonglines [LLD]	0	2	1	0	0	3	0	0	0	0	1	2	0
Set longlines [LLS]	3	4	3	1	0	4	1	0	3	2	3	4	3
Pots and Traps [FPO]	0	4	4	0	0	4	4	2	5	4	4	3	4
Fykenets [FYK]	0	3	0	0	0	0	0	0	1	3	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	4	3	0	0	5	1	0	4	3	3	2	1
Set gillnet [GNS]	3	4	3	2	1	5	3	0	4	3	4	3	3
Driftnet [GND]	0	3	1	0	0	0	0	0	3	1	2	0	0
Purse-seine [PS]	3	3	1	2	0	4	1	0	2	1	2	1	1
Lampara nets [LA]	0	0	0	0	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	2	0	0	0	2	0	3	1	2	2	2
Anchored seine [SDN]	0	0	0	0	0	0	1	0	3	3	1	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach-seine [SB]	0	0	1	0	0	0	0	0	0	2	1	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.1. Identified risk factor for the species groups for each métier in the Azores

MÉTIER	AZ								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	0	0	0	0	0	0	0	0	0
Bottom otter trawl [OTB]	0	0	0	0	0	0	0	0	0
Multi-rig otter trawl [OTT]	0	0	0	0	0	0	0	0	0
Bottom pair trawl [PTB]	0	0	0	0	0	0	0	0	0
Beam trawl [TBB]	0	0	0	0	0	0	0	0	0
Midwater otter trawl [OTM]	0	0	0	0	0	0	0	0	0
Pelagic pair trawl [PTM]	0	0	0	0	0	0	0	0	0
Hand and Pole lines [LHP] [LHM]	4	4	4	4	4	4	4	0	4
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	3	3	9	6	9	3	3	0	6
Pots and Traps [FPO]	0	0	0	0	0	0	0	0	0
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	0	0	0	0	0	0	0	0
Set gillnet [GNS]	3	9	9	9	3	9	6	0	6
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	3	3	3	3	3	3	6	0	3
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.2. Identified risk factor for the species groups for each métier in the Bay of Biscay

MÉTIER	BB								
	lamprey s	round fish	turtle s	diving birds	surface birds	seal s	dolphin s	harbour porpoise	large whales
Boat dredge [DRB]	3	3	3	3	3	3	3	3	3
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	8	8	12	4	4	4	4	4	4
Bottom pair trawl [PTB]	6	6	9	3	3	3	3	3	3
Beam trawl [TBB]	4	2	6	2	2	2	2	2	2
Midwater otter trawl [OTM]	3	9	6	6	6	6	6	3	6
Pelagic pair trawl [PTM]	3	9	6	6	6	6	6	3	6
Hand and Pole lines [LHP] [LHM]	3	3	3	3	3	3	3	3	3
Trolling lines [LTL]	3	3	3	6	9	3	3	3	3
Drifting longlines [LLD]	2	2	6	4	6	2	2	2	4
Set longlines [LLS]	4	4	12	8	12	4	4	4	8
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	9	6	3	6	3	9	3	3	3
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	4	12	12	12	4	12	8	12	8
Set gillnet [GNS]	4	12	12	12	4	12	8	12	8
Driftnet [GND]	3	9	9	9	9	9	9	9	9
Purse-seine [PS]	3	3	3	3	3	3	6	3	3
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.3. Identified risk factor for the species groups for each métier in the Celtic Sea

MÉTIER	CS								
	lamprey s	round fish	turtle s	diving birds	surface birds	seal s	dolphin s	harbour porpoise	large whales
Boat dredge [DRB]	2	2	2	2	2	2	2	2	2
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	8	8	12	4	4	4	4	4	4
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	6	3	9	3	3	3	3	3	3
Midwater otter trawl [OTM]	2	6	4	4	4	4	4	2	4
Pelagic pair trawl [PTM]	2	6	4	4	4	4	4	2	4
Hand and Pole lines [LHP] [LHM]	3	3	3	3	3	3	3	3	3
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	3	3	9	6	9	3	3	3	6
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	3	9	9	9	3	9	6	9	6
Set gillnet [GNS]	3	9	9	9	3	9	6	9	6
Driftnet [GND]	1	3	3	3	3	3	3	3	3
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	4	4	2	2	2	2	2	2	2
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.4. Identified risk factor for the species groups for each métier in the Eastern Artic

MÉTIER	EA								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	0	0	0	0	0	0	0	0	0
Bottom otter trawl [OTB]	6	6	0	3	6	6	3	3	3
Multi-rig otter trawl [OTT]	0	0	0	0	0	0	0	0	0
Bottom pair trawl [PTB]	4	4	0	2	2	2	2	2	2
Beam trawl [TBB]	0	0	0	0	0	0	0	0	0
Midwater otter trawl [OTM]	3	9	0	6	6	6	6	3	6
Pelagic pair trawl [PTM]	1	3	0	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	0	0	0	0	0	0	0	0	0
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	1	1	0	2	3	1	1	1	2
Pots and Traps [FPO]	0	0	0	0	0	0	0	0	0
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	0	0	0	0	0	0	0	0
Set gillnet [GNS]	2	6	0	6	2	6	4	6	4
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	2	2	0	2	2	2	4	2	2
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.5. Identified risk factor for the species groups for each métier in the Faroe Island

MÉTIER	FI								
	lampre ys	round fish	turtl es	diving birds	surface birds	seal s	dolphi ns	harbour porpoise	large whales
Boat dredge [DRB]	0	0	0	0	0	0	0	0	0
Bottom otter trawl [OTB]	4	4	6	2	4	4	2	2	2
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	0	0	0	0	0	0	0	0	0
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	0	0	0	0	0	0	0	0	0
Hand and Pole lines [LHP] [LHM]	0	0	0	0	0	0	0	0	0
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	0	0	0	0	0	0	0	0	0
Pots and Traps [FPO]	0	0	0	0	0	0	0	0	0
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	0	0	0	0	0	0	0	0
Set gillnet [GNS]	1	3	3	3	1	3	2	3	2
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	0	0	0	0	0	0	0	0	0
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.6. Identified risk factor for the species groups for each métier in the Iberian Waters

MÉTIER	IB								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	4	4	4	4	4	4	4	4	4
Bottom otter trawl [OTB]	10	10	15	5	10	10	5	5	5
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	8	8	12	4	4	4	4	4	4
Beam trawl [TBB]	8	4	12	4	4	4	4	4	4
Midwater otter trawl [OTM]	1	3	2	2	2	2	2	1	2
Pelagic pair trawl [PTM]	2	6	4	4	4	4	4	2	4
Hand and Pole lines [LHP] [LHM]	4	4	4	4	4	4	4	4	4
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	3	3	9	6	9	3	3	3	6
Set longlines [LLS]	4	4	12	8	12	4	4	4	8
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	5	15	15	15	5	15	10	15	10
Set gillnet [GNS]	5	15	15	15	5	15	10	15	10
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	4	4	4	4	4	4	8	4	4
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.7. Identified risk factor for the species groups for each métier in the Irish Sea

MÉTIER	IS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	3	3	3	3	3	3	3	3	3
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	2	2	3	1	1	1	1	1	1
Bottom pair trawl [PTB]	0	0	0	0	0	0	0	0	0
Beam trawl [TBB]	6	3	9	3	3	3	3	3	3
Midwater otter trawl [OTM]	2	6	4	4	4	4	4	2	4
Pelagic pair trawl [PTM]	2	6	4	4	4	4	4	2	4
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	1	1	3	2	3	1	1	1	2
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	3	9	9	9	3	9	6	9	6
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	4	4	2	2	2	2	2	2	2
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.8. Identified risk factor for the species groups for each métier in the Mid-Atlantic

MÉTIER	MA								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	0	0	0	0	0	0	0	0	0
Bottom otter trawl [OTB]	6	6	9	3	6	6	3	3	3
Multi-rig otter trawl [OTT]	0	0	0	0	0	0	0	0	0
Bottom pair trawl [PTB]	0	0	0	0	0	0	0	0	0
Beam trawl [TBB]	0	0	0	0	0	0	0	0	0
Midwater otter trawl [OTM]	2	6	4	4	4	4	4	2	4
Pelagic pair trawl [PTM]	1	3	2	2	2	2	2	1	2
Hand and Pole lines [LHP] [LHM]	0	0	0	0	0	0	0	0	0
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	0	0	0	0	0	0	0	0	0
Pots and Traps [FPO]	4	2	2	2	2	2	2	2	4
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	0	0	0	0	0	0	0	0
Set gillnet [GNS]	0	0	0	0	0	0	0	0	0
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	0	0	0	0	0	0	0	0	0
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	0	0	0	0	0	0
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.9. Identified risk factor for the species groups for each métier in the North Sea and Eastern Channel

MÉTIER	NS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	4	4	0	4	4	4	4	4	4
Bottom otter trawl [OTB]	10	10	0	5	10	10	5	5	5
Multi-rig otter trawl [OTT]	8	8	0	4	4	4	4	4	4
Bottom pair trawl [PTB]	6	6	0	3	3	3	3	3	3
Beam trawl [TBB]	8	4	0	4	4	4	4	4	4
Midwater otter trawl [OTM]	3	9	0	6	6	6	6	3	6
Pelagic pair trawl [PTM]	3	9	0	6	6	6	6	3	6
Hand and Pole lines [LHP] [LHM]	3	3	0	3	3	3	3	3	3
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	3	3	0	6	9	3	3	3	6
Pots and Traps [FPO]	10	5	0	5	5	5	5	5	10
Fykenets [FYK]	3	2	0	2	1	3	1	1	1
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	4	12	0	12	4	12	8	12	8
Set gillnet [GNS]	4	12	0	12	4	12	8	12	8
Driftnet [GND]	3	9	0	9	9	9	9	9	9
Purse-seine [PS]	2	2	0	2	2	2	4	2	2
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	6	6	0	3	3	3	3	3	3
Anchored seine [SDN]	6	6	0	3	3	3	3	3	3
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.10. Identified risk factor for the species groups for each métier in the Skagerrak and Kattegat

MÉTIER	SK								
	lampreys	**round fish**	**turtles**	**diving birds**	**surface birds**	**seals**	**dolphins**	**harbour porpoise**	**large whales**
Boat dredge [DRB]	2	2	0	2	2	2	2	2	2
Bottom otter trawl [OTB]	8	8	0	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	6	6	0	3	3	3	3	3	3
Bottom pair trawl [PTB]	4	4	0	2	2	2	2	2	2
Beam trawl [TBB]	4	2	0	2	2	2	2	2	2
Midwater otter trawl [OTM]	2	6	0	4	4	4	4	2	4
Pelagic pair trawl [PTM]	2	6	0	4	4	4	4	2	4
Hand and Pole lines [LHP] [LHM]	3	3	0	3	3	3	3	3	3
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	2	2	0	4	6	2	2	2	4
Pots and Traps [FPO]	8	4	0	4	4	4	4	4	8
Fykenets [FYK]	9	6	0	6	3	9	3	3	3
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	3	9	0	9	3	9	6	9	6
Set gillnet [GNS]	3	9	0	9	3	9	6	9	6
Driftnet [GND]	1	3	0	3	3	3	3	3	3
Purse-seine [PS]	1	1	0	1	1	1	2	1	1
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	2	2	0	1	1	1	1	1	1
Anchored seine [SDN]	6	6	0	3	3	3	3	3	3
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	4	4	0	2	2	2	2	2	2
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.11. Identified risk factor for the species groups for each métier in the Western Channel

MÉTIER	WC								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	4	4	4	4	4	4	4	4	4
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	4	4	6	2	2	2	2	2	2
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	6	3	9	3	3	3	3	3	3
Midwater otter trawl [OTM]	3	9	6	6	6	6	6	3	6
Pelagic pair trawl [PTM]	3	9	6	6	6	6	6	3	6
Hand and Pole lines [LHP] [LHM]	3	3	3	3	3	3	3	3	3
Trolling lines [LTL]	2	2	2	4	6	2	2	2	2
Drifting longlines [LLD]	1	1	3	2	3	1	1	1	2
Set longlines [LLS]	3	3	9	6	9	3	3	3	6
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	3	9	9	9	3	9	6	9	6
Set gillnet [GNS]	4	12	12	12	4	12	8	12	8
Driftnet [GND]	2	6	6	6	6	6	6	6	6
Purse-seine [PS]	2	2	2	2	2	2	4	2	2
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	4	4	2	2	2	2	2	2	2
Anchored seine [SDN]	2	2	1	1	1	1	1	1	1
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	2	2	1	1	1	1	1	1	1
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.12. Identified risk factor for the species groups for each métier in the Western Ireland

MÉTIER	WI								
	lampreys	**round fish**	**turtles**	**diving birds**	**surface birds**	**seals**	**dolphins**	**harbour porpoise**	**large whales**
Boat dredge [DRB]	2	2	2	2	2	2	2	2	2
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	6	6	9	3	3	3	3	3	3
Bottom pair trawl [PTB]	2	2	3	1	1	1	1	1	1
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	3	9	6	6	6	6	6	3	6
Pelagic pair trawl [PTM]	3	9	6	6	6	6	6	3	6
Hand and Pole lines [LHP] [LHM]	2	2	2	2	2	2	2	2	2
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	2	2	6	4	6	2	2	2	4
Set longlines [LLS]	4	4	12	8	12	4	4	4	8
Pots and Traps [FPO]	6	3	3	3	3	3	3	3	6
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	2	6	6	6	2	6	4	6	4
Set gillnet [GNS]	3	9	9	9	3	9	6	9	6
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	4	4	2	2	2	2	2	2	2
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6.13. Identified risk factor for the species groups for each métier in the Western Scotland

MÉTIER	WS								
	lampreys	round fish	turtles	diving birds	surface birds	seals	dolphins	harbour porpoise	large whales
Boat dredge [DRB]	3	3	3	3	3	3	3	3	3
Bottom otter trawl [OTB]	8	8	12	4	8	8	4	4	4
Multi-rig otter trawl [OTT]	6	6	9	3	3	3	3	3	3
Bottom pair trawl [PTB]	4	4	6	2	2	2	2	2	2
Beam trawl [TBB]	2	1	3	1	1	1	1	1	1
Midwater otter trawl [OTM]	3	9	6	6	6	6	6	3	6
Pelagic pair trawl [PTM]	2	6	4	4	4	4	4	2	4
Hand and Pole lines [LHP] [LHM]	1	1	1	1	1	1	1	1	1
Trolling lines [LTL]	0	0	0	0	0	0	0	0	0
Drifting longlines [LLD]	0	0	0	0	0	0	0	0	0
Set longlines [LLS]	3	3	9	6	9	3	3	3	6
Pots and Traps [FPO]	8	4	4	4	4	4	4	4	8
Fykenets [FYK]	0	0	0	0	0	0	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	1	3	3	3	1	3	2	3	2
Set gillnet [GNS]	3	9	9	9	3	9	6	9	6
Driftnet [GND]	0	0	0	0	0	0	0	0	0
Purse-seine [PS]	1	1	1	1	1	1	2	1	1
Lampara nets [LA]	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	4	4	2	2	2	2	2	2	2
Anchored seine [SDN]	0	0	0	0	0	0	0	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 7.Summed risk factors for each métier at different areas (summed across all species)

	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	0	27	18	0	0	36	27	0	32	16	36	18	27
Bottom otter trawl [OTB]	0	60	60	36	30	75	60	45	60	48	60	60	60
Multi-rig otter trawl [OTT]	0	52	52	0	13	13	13	0	40	30	26	39	39
Bottom pair trawl [PTB]	0	39	13	20	13	52	0	0	30	20	13	13	26
Beam trawl [TBB]	0	24	36	0	0	48	36	0	36	18	36	12	12
Midwater otter trawl [OTM]	0	51	34	45	17	17	34	34	45	30	51	51	51
Pelagic pair trawl [PTM]	0	51	34	15	0	34	34	17	45	30	51	51	34
Hand and Pole lines [LHP] [LHM]	32	27	27	0	0	36	9	0	24	24	27	18	9
Trolling lines [LTL]	0	36	0	0	0	0	0	0	0	0	24	0	0
Drifting longlines [LLD]	0	30	15	0	0	45	0	0	0	0	15	30	0
Set longlines [LLS]	42	60	45	12	0	60	15	0	36	24	45	60	45
Pots and Traps [FPO]	0	44	44	0	0	44	44	22	50	40	44	33	44
Fykenets [FYK]	0	45	0	0	0	0	0	0	14	42	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	84	63	0	0	105	21	0	72	54	63	42	21
Set gillnet [GNS]	54	84	63	36	21	105	63	0	72	54	84	63	63
Driftnet [GND]	0	75	25	0	0	0	0	0	66	22	50	0	0
Purse-seine [PS]	27	30	10	18	0	40	10	0	18	9	20	10	10
Lampara nets [LA]	0	0	0	0	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	22	0	0	0	22	0	30	10	22	22	22
Anchored seine [SDN]	0	0	0	0	0	0	11	0	30	30	11	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	11	0	0	0	0	0	0	20	11	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 8.Summed risk factors for each species at different areas (summed across all metiers)

	LAMPR EYS	ROUND FISH	TURTLES	DIVING BIRDS	SURFACE BIRDS	SEALS	DOLPHINS	HARBOUR PORPOISE	LARGE WHALES
AZ	13	19	25	22	19	19	19	0	19
BB	78	103	121	95	89	93	78	77	85
CS	59	74	90	62	58	62	53	54	60
EA	19	31	0	23	23	25	22	18	21
FI	10	14	17	9	9	11	8	8	8
IB	68	86	115	80	72	78	67	70	74
IS	44	53	59	42	39	45	38	37	42
MA	13	17	17	11	14	14	11	8	13
NS	86	110	0	89	80	92	79	79	85
SK	70	83	0	66	57	71	56	57	61
WC	66	89	100	78	74	76	67	66	73
WI	50	68	82	58	58	56	48	46	56
WS	49	62	72	49	48	50	43	41	49

Table 9. Number of trips sampled on-board by métier for each fishing ground under the DCF as reported by MS National Plans 2011-2013

MÉTIER	no of trips	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	70			7				31		20		12		
Bottom otter trawl [OTB]	4,149		134	113	17	32	652	266	13	421	425	127	97	145
Multi-rig otter trawl [OTT]	12									12				
Bottom pair trawl [PTB]	183		24		6		141							
Beam trawl [TBB]	316			40			36	6		161		72	1	
Midwater otter trawl [OTM]	524		40		15		4	36	6	62		12	148	18
Pelagic pair trawl [PTM]	125		13	10	1		8	16		6		8	24	9
Hand and Pole lines [LHP] [LHM]	180	144	0	0	0	0	0	0	0	0	0	0	0	0
Trolling lines [LTL]	0													
Driftinglonglines [LLD]	268													
Set longlines [LLS]	252	144	36				36			0				
Pots and Traps [FPO]	384			11			0	202		12	36	0	11	34
Fykenets [FYK]	33									0	30			
Stationary uncovered poundnets [FPN]	102													
Trammelnet [GTR]	432		120	6			40			40		46		
Set gillnet [GNS]	711			56		7	126	8	7	105	60	56	16	2
Driftnet [GND]	0													
Purse-seine [PS]	270						96							
Lampara nets [LA]	15													
Fly shooting seine [SSC]	44			2				3		36			3	
Anchored seine [SDN]	60		12							12	36			
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach-seine [SB]	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 10. Relative DCF sampling effort (table 9) subtracted from relative summed risk factors (table 7) for each métier at different areas. Positive numbers (in green), indicate relative under sampling; negative numbers (in red) indicate relative over sampling.

	AZ	BB	CS	EA	FI	IB	IS	MA	NS	SK	WC	WI	WS
Boat dredge [DRB]	0	3	0	0	0	5	1	0	2	3	2	3	6
Bottom otter trawl [OTB]	0	-28	-36	-24	-50	-47	-32	-12	-39	-63	-29	-21	-57
Multi-rig otter trawl [OTT]	0	6	9	0	14	2	3	0	4	6	4	7	8
Bottom pair trawl [PTB]	0	-2	2	-4	14	-5	0	0	4	4	2	2	6
Beam trawl [TBB]	0	3	-10	0	0	4	8	0	-13	3	-16	2	3
Midwater otter trawl [OTM]	0	-4	6	-14	18	2	2	6	-1	6	4	-40	2
Pelagic pair trawl [PTM]	0	3	2	6	0	4	6	14	6	6	5	2	3
Hand and Pole lines [LHP] [LHM]	-29	3	5	0	0	5	2	0	3	5	4	3	2
Trolling lines [LTL]	0	4	0	0	0	0	0	0	0	0	3	0	0
Drifting longlines [LLD]	0	4	3	0	0	6	0	0	0	0	2	6	0
Set longlines [LLS]	-23	-2	8	7	0	5	4	0	5	5	7	11	10
Pots and Traps [FPO]	0	5	3	0	0	6	-25	19	6	2	6	3	-7
Fykenets [FYK]	0	5	0	0	0	0	0	0	2	3	0	0	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0	0	0	0	0	0	0	0
Trammelnet [GTR]	0	-21	9	0	0	11	5	0	6	10	-5	8	5
Set gillnet [GNS]	35	10	-12	20	4	4	14	-27	-2	0	-5	7	13
Driftnet [GND]	0	9	4	0	0	0	0	0	9	4	7	0	0
Purse-seine [PS]	17	4	2	10	0	-3	3	0	3	2	3	2	2
Lampara nets [LA]	0	0	0	0	0	0	0	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	3	0	0	0	5	0	0	2	3	3	5
Anchored seine [SDN]	0	-3	0	0	0	0	3	0	3	0	2	0	0
Pair seine [SPR]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	2	0	0	0	0	0	0	4	2	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 11. Summary table presenting for each métier at the Bay of Biscay and the North Sea and Eastern Channel: the summed risk factors (table 7), a categorization of these risk factors (0:0, 1: 1–25, 2: 26–50, 3: 51–75, 4: 76–150), and the DCF sampling effort (table 9).

	BB			NS		
	SUM RISK	RISK CAT	DCF SAMP	SUM RISK	RISK CAT	DCF SAMP
Boat dredge [DRB]	27	2	0	32	2	20
Bottom otter trawl [OTB]	60	3	134	60	3	421
Multi-rig otter trawl [OTT]	52	3	0	40	2	12
Bottom pair trawl [PTB]	39	2	24	30	2	0
Beam trawl [TBB]	24	1	0	36	2	161
Midwater otter trawl [OTM]	51	3	40	45	2	62
Pelagic pair trawl [PTM]	51	3	13	45	2	6
Hand and Pole lines [LHP] [LHM]	27	2	0	24	1	0
Trolling lines [LTL]	36	2	0	0	0	0
Drifting longlines [LLD]	30	2	0	0	0	0
Set longlines [LLS]	60	3	36	36	2	0
Pots and Traps [FPO]	44	2	0	50	2	12
Fykenets [FYK]	45	2	0	14	1	0
Stationary uncovered poundnets [FPN]	0	0	0	0	0	0
Trammelnet [GTR]	84	4	120	72	3	40
Set gillnet [GNS]	84	4	0	72	3	105
Driftnet [GND]	75	3	0	66	3	0
Purse-seine [PS]	30	2	0	18	1	0
Lampara nets [LA]	0	0	0	0	0	0
Fly shooting seine [SSC]	0	0	0	30	2	36
Anchored seine [SDN]	0	0	12	30	2	12
Pair seine [SPR]	NA	NA	NA	NA	NA	NA
Beach and boat seine [SB] [SV]	0	0	0	0	0	0
Glass eel fishing	NA	NA	NA	NA	NA	NA

ANNEX 16 WP3 A regional sampling plan for the collection of stomach sampling data

The fishPi project
Regional co-operation in fisheries data collection

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP3 – Regional sampling programmes for fisheries and ecosystem impact data not currently collected (including stomach contents and by-catch)

Deliverable 3.2 A regional sampling plan for data collection of stomach data

Lead: AZTI, Spain

Core Team: Estanis Mugerza, Lucia Zarauz and Eider Andonegi (AZTI, Spain)
Hugo Mendes and Ana Moreno (IPMA, Portugal)
Mike Armstrong, (CEFAS, UK)
Patrik Börjesson (SLU, Sweden)

Table of Contents

1 A regional sampling plan for data collection of stomach contents

1 Introduction

The support of polices for sustainable management strategies of living marine resources, demands integrated ecosystem advice on the long-term impacts of fisheries. An improved regional sampling of relevant variables would improve the understanding of fisheries impacts on the ecosystem and align the Data Collection Multi-Annual Plan(DC-MAP) with obligations under other existing EU legislative instruments. In 2008, the EU published the Marine Strategy Framework Directive (MSFD, EU, 2008) establishing a framework for Member States (MS) to take the necessary measures to achieve or maintain Good Environmental Status (GES) in the marine environment by 2020. The MSFD also highlights the relevance of the ecosystem approach (EA) for ensuring the conservation of marine systems. The MSFD is defined by eleven qualitative descriptors, one of which (D4) focuses explicitly on food webs.

More recently, the reform of the Common Fisheries Policy (CFP; EU, 2013) anticipates more extensive use of long-term management plans and refers to the Ecosystem Approach to Fisheries (EAF) or Ecosystem Based Fisheries Management (EBAF) as a means to restore stocks to levels that can produce the Maximum Sustainable Yield (MSY). Both long term management plans and estimates of the fishing mortality providing MSY are particularly sensitive to changes in natural mortality, and a prerequisite for correctly estimating natural mortality is the accurate knowledge of species interactions (Huwer et al., 2014). In particular, the CFP highlights the importance of the existing trophic relations between species and the need for including specific alternative conservation measures based on ecosystem, for an appropriate assessment of fish species. Food-webs have become a major focus for EU research and maritime policy, yet there are surprisingly few long-term datasets available within a European context for parameterizing models of predator-prey interactions in the ocean. There is growing demand for data on “who eats whom” in marine systems, in order to deduce how changes in one part of the ecosystem might have consequences elsewhere (Pinnegar, 2014b). Additionally, ICES have stated that stomach data are of vital importance and that it intends to gradually transition to providing multispecies advice on fisheries for European ecosystems in the near future (ICES, 2013a).

1.1 Justification of data collection

While traditional fisheries management focuses on harvest rates and stock biomass, incorporating the impacts of ecosystem processes are one of the main pillars of the EAF. Although EAF has been formally adopted widely since the 1990s, the implementation is largely a bottom-up process mostly driven by a few dedicated experts, and tactical fisheries management is still predominantly single-species oriented taking little account of ecosystem processes. Thus, while the ecosystem approach is highlighted in policy, key aspects of it tend not to be implemented in actual fisheries management (Skern-Mauritze et al., 2016). Hence, information on prey availability, competition, predation processes or biotoxins and plastic particles levels in fish stomachs are needed to support several policies (e.g. CFP, MSFD) that envisage an EAF and an EBFM. Assessing trophic relations with detailed stomach contents analysis increases knowledge on suitable stock-recruit models (e.g. density dependent effects like cannibalism), assessment of fish species (e.g. estimates of Natural Mortality), reliable Biological Reference Points (BRP) considering species interaction, all aiming at providing a more appropriate framework for the implementation of multi-annual management plans. New data on predation is also important for providing both tactical and strategic advice for management of marine ecosystems (FAO, 2008), since they positively contribute to the quality of the tools used to quantitatively assess their dynamics (i.e. multispecies assessment models, ecosystems models, etc.).

Also data on marine litter (plastic) and biotoxins levels in the environment could be collected from the analysis of stomach contents as the trophic transfer of contaminants throughout the food web can have serious consequences on aquaculture, fisheries, human health, industries and economies. Consequently this information will help better understanding the potential effects of existing anthropogenic and natural stressors on the natural integrity of the ecosystems and so, on the human well-being to some extent (ICES, 2016). In other words, using good data that reflect existing trophic relationships would support the scientific community to move towards implementing Integrated Ecosystem Assessments (IEA) (Levin et al., 2009), following the approach provided by ICES aiming at providing appropriate scientific advice for an Ecosystem Based Management (EBM) of marine systems (Dickey-Collas, 2014; Walther and Möllmann, 2014)

The information currently collected under the Data Collection Framework (DCF; EU, 2008) seems not to be enough to achieve these goals, and new requirements have been proposed by end users (Table 20). These requirements include the collection of stomach contents and although, stomach sampling is not covered by the current DCF, some MS already collect this information. But it is not a general practice and in most cases, the sampling is not coordinated at a regional scale and the information obtained might not be available for the scientific community.

1.2 Decision making

There are a wide variety of potential end users in Europe that require or might require the use of stomach content data, most of them having an important role within the decision making process (Table 20).

Table 13. Existing and potential end-users of stomach data.

End User	End user sub groups	Use of stomach data	Role in decision making process
ICES	Working Group on Multispecies Assessment Methods (WGSAM)	Multispecies and ecosystem models, including EwE, SMS, Gadget models, used to provide natural mortality for single species assessments	Increasing in the last years, depending on the region
	Working Group on Ecosystem Effects of Fishing Activities (WGECO)	Specific examples of fisheries impacts in the ecosystem, MSFD descriptors	Potentially YES
	Regional Integrated Assessment Working Groups (WGINOSE, WGNARS, WGEAWESS, WGIAB, WGIBAR, WGINOR, WGCOMEDA)	Ecosystem models (EwE, Atlantis), multispecies models (SMS, Gadget) – strongly focused on biodiversity, climate change issues, large marine ecosystem functioning, and the context of ecosystem health indicators development for the Marine Strategy Framework Directive.	Increasing in the last years, depending on the region
	Steering group on integrated ecosystem assessment	Support integrated ecosystem assessments	Increasing in the last years, depending on the region
	Steering group on ecosystem pressures and impacts	Support integrated ecosystem assessments	Increasing in the last years, depending on the region
	Steering group on ecosystem dynamics and processes	Support integrated ecosystem assessments	Increasing in the last years,

			depending on the region
	SCICOM	Support integrated ecosystem assessments	YES, increasing in the last years
	Stock assessment working groups	Estimates on natural mortality	YES
	ACOM	Advice for management based on Stock assessments	YES
European Commission	DGMARE & DG Environment	Implementation of MSFD; achievement of GES with good management of recreational as well as commercial fishery impacts. Implementation of CE 812/2004, Birds Directive, Habitats Directive	YES
	STECF	Inclusion of data collection in the EU MAP	YES
International Organisatons	FAO, OSPAR, ASCOBANS, ACAP, IWC, HELCOM	Identifying threats, recommending action plans, implementation of different agreements	YES
Regional Coordination Groups	RCGs for each region	Coordination and cost-effectiveness of by catch data collection within regions (if included in EU-MAP)	YES
National Governments and regional fisheries authorities within countries	Commercial fishermen's organisations and federations.	Developing policy positions on management that reflects the ecosystem aspects of sustainable development in coastal regions and spatial planning such as MCZs. Meeting international agreed responsibilities	YES
Scientific community in general.	Universities; Govt. departments; other Institutes	Scientists interested in how food webs and ecosystem dynamics are changing in each region and occurrences of species beyond previous range. Data for publication	Potentially
Representative bodies for International and national commercial fisheries.		Policy developments in relation to interaction between commercial species and main predators;	YES
Advisory Councils	e.g. North Western Waters AC; North Sea AC…..	Policy developments in relation to interaction between commercial species and main predators	YES
Marine NGOs		Policy developments in relation to interaction between commercial species and main predators	YES

1.3 Regional objectives and estimates

1.3.1 Identify different estimates needs

Stomach content data serve as the basis for a range of multispecies, extended single-species, and ecosystem models where the main focus is on who-eats-who and how much. Having a solid foundation of adequate stomach content data are a prerequisite for implementing the EAF and EBFM, but most of the studies on the diet of predators are based on samples of one or two years, and it is not possible to determine whether these years reflect the average feeding behaviour or were incidental years of a given species consumption. This concern is aggravated by the knowledge of the major changes in European waters occurring in the last 20 years, including shifts in benthic, plankton and fish communities. One aspect of these changes, mostly explained by an increase in temperature observed over the last decade in the eastern North Atlantic, is that the distributions of several northern fish stocks have changed with many of the cold water species shifting their distributions towards higher latitudes and more southern and subtropical species increasing in abundance. With significant ecological effects, these shifts have resulted in heterogeneous changes in fish abundance and have significant consequences for fisheries and their management (Brander et al., 2003; Stebbing et al., 2002; Perry et al., 2005; Last et al., 2011, Serra et al., 2015). While adaptive measures have been suggested from both resource assessment and management perspectives, the only assumption is that the change in abundance is the only parameter affected. However, changes in marine communities structure also changes the trophodynamic relationships (spatial and temporal diet variability) and induces changes in the mix of dominating species over decadal time scales. Trophic characteristics may favour the dominance of one fish species over another when suitable oceanographic and ecological conditions for the development of the appropriate food occur. As food webs are fully interconnected systems, these changes on one part of the system may have impacts elsewhere – by top-down or bottom-up control – that are not easily predictable.

For all these caveats and because of the limited trust in the stability of trophodynamics, management target fishing mortalities and reference points are still estimated not accounting for trophic interactions or at most, assuming that predation is constant over the whole stock of prey and disregarding prey availability effects on growth and recruitment. ICES demonstrated the sensitivity of the North Sea (NS) cod stock assessment to multispecies predation models. Cod is a top predator in the NS ecosystem with important commercial prey fish as whiting, haddock, herring and cod itself. Cannibalism lead to relatively high MSY reference points values when estimated in a multispecies context compared to single species approaches. Predictions also showed that fishing cod at too low fishing mortalities may have negative consequences for the stock dynamics of whiting and haddock. These results were hampered by the lack of recent stomach content data and the low number of stomachs sampled in some of the NS areas (ICES, 2012).

A comprehensive investigation of species interactions is needed on a regional basis. It is clear that more detailed food web information will be needed to improve predictions of the response of stocks to ongoing perturbations which are amplified by the lack of comprehensive long-term datasets, and lack of spatial and temporal data collection. To move to the point where all of our existing knowledge of interactions can be used as basis for long term management, stomach data covering regional areas and preferably in a long time series are needed.

1.3.2 Marine Strategy Framework Directive

The goal of the MSFD is the protection and conservation of Europe marine environment to achieve GES by 2020. The strategy requires MS to assess the environmental status of their marine waters, and to determine what is considered to be GES. MS are also required to work together at a regional or sub-regional level to ensure a consistent determination of good environmental every six years. This means that MS will have to update their initial assessments, their determination of GES and their environmental targets in 2018, when the second cycle of implementation of the MSFD will start.

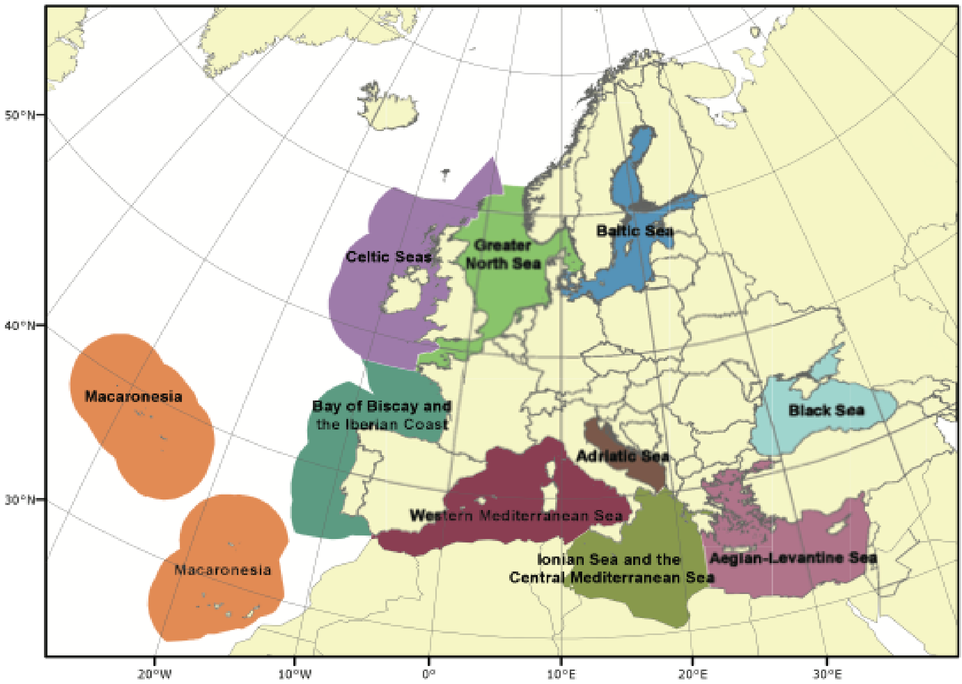


Figure 7. MSFD European Marine Regions on the basis of geographical and environmental criteria.

The food web descriptor is generally regarded as the most challenging descriptor. The Commission agreed that the limited state of advancement and knowledge of D4 criteria, indicators and reference levels, is expected to need further development in order to be operational and to dictate more explicit monitoring needs. Both Joint Research Centre (JRC) and ICES recommend further developmental work to underpin the implementation of D4 and agreed that data on species diet by stomach sampling and stable isotope analysis (see later section 4) should support this descriptor and could be very useful to estimate species trophic level and thus establish suitable indicators.

The Directive definition of GES is rather high level and a revised decision will be made available during 2016 listing updated objectives, notably, revised criteria and methodological standards for biodiversity (D1), food webs (D4) and sea-floor integrity (D6), among others. It includes criteria to be used in relation to the assessment of ecosystem state characteristics and that assessment of D1, D4 and D6, should be split into the following themes: i) birds, mammals, reptiles, fish and cephalopods, ii) pelagic and benthic habitats and iii) ecosystems, including food webs.

From the revised decision, D4 assessment now includes the biomass of, and size distribution within of key trophic guilds and the assessment of species diversity has been coupled to food web criteria, to respect both the need to assess biodiversity under D1 and to assess diversity under D4. The recommendation is that each MS under the MSFD should establish a monitoring program (or use current knowledge and programs) to assess a set of regional representative guilds. Not all trophic guilds in each ecosystem need to be assessed but that, by region, a minimum of at least three trophic guilds should be monitored, preferably covering both lower and higher trophic levels (Table 14). Moreover, improved regional coherence is sought through development of reference levels at (sub)regional level and assessment at appropriate ecological scales as shown in Figure 7, rather than using national boundaries. (EU, 2016).

Table 14. Trophic guild examples. X denotes where the taxonomic groups contribute significantly to each guild. Nekton includes bony fish, elasmobranchs, and squids (adapted ICES, 2015b).

Guild/Species	Phytoplankton	Zooplankton	Benthos	Nekton excl.	Seabirds	Marine

group				warm blooded		mammals
Primary producers	x					
Secondary producers		x				
Filter-feeders			x			
Deposit-feeders			x			
Planktivores			x	x	x	x
Sub-apex pelagic predators				x	x	x
Sub-apex demersal predators			x	x	x	x
Apex predators				x	x	x

Additional food web indicators have been proposed and can also be potentially developed from stomach information and isotopic analysis. Extensive comprehensive guidance on their application and efficiency has been provided by several ICES study groups (for details see ICES, 2013b; 2014) and also within the OSPAR Commission and JRC.

A monitoring program under MSFD D10 (Marine litter) is also relevant and should be taken into consideration in a stomach collection protocol. Many marine organisms are known to ingest marine litter and litter has been found in stomachs of sea birds, marine mammals and fish. A stomach sampling programme can also provide stomach or tissue samples of susceptible species for the analysis of ingestion rates of micro- and nanoplastics. The monitoring of microplastics may be closely related to monitoring of contaminants under D8 (Contaminants) and D9 (Contaminants in seafood), as microplastics are known to be adhesive to pollutants. Recently, by request of OSPAR, ICES provided a preliminary protocol for monitoring of plastics in fish stomachs that enables the analysis of fish stomachs for plastic particles in the marine environment to establish a baseline with which to gauge whether or not trends in plastic are increasing or decreasing in the future, leading to possible management actions. This protocol was recommended to be reviewed at regular intervals and improved on the basis of experience (ICES, 2015a).

There has been some discussion that biotoxins are not expressly included under MSFD Descriptor 8 (Contaminants) or 9 (Contaminants in seafood) as the threat from marine biotoxins is already managed in a different manner with established regulatory levels. Regardless, a stomach collection protocol can also guide monitoring activities for D5 (Eutrophication) and surveillance of marine biotoxins. The trophic transfer of biotoxins throughout the food web can have serious consequences on aquaculture, fisheries, human health, industries and economies. Both fish and shellfish have two main roles in the food web regarding toxins exposure and impact; they may act as vectors of toxins, as victims or as both. Due to the implications of toxins to human health, the role of shellfish is well documented, leading most coastal countries to conduct monitoring programmes for marine toxins. Contrary, as fish are not the target for toxins monitoring, data of bio accumulated toxins in fish are much less available Although much attention has been paid to the effects of phycotoxins in higher organisms including marine mammals, sea birds and also humans, the trophic links between these organisms should also be elucidated. In Europe, extensive research has focussed on biotoxic accumulation in bivalves; however, studies of toxin transfer through the pelagic food chain are scarce (Costa and Garrido, 2004). During a harmful toxin event, high cell densities of toxin-producing phytoplankton species are attained and are therefore easily ingested by planktivorous fish. Some small pelagic fishes (sardines, anchovies) have been proposed to be particularly relevant since they are able to filter and digest phytoplankton species efficiently and can accumulate large quantities of toxins in their guts and several studies indicate that planktivorous fish are potentially more effective toxin vectors than bivalves (Lefebvre et al. 2002, Costa and Garrido, 2004) .

The knowledge obtained from a stomach collection protocol can potentially be used to assess and guide monitoring activities under MSFD D1, D5, D10 (Table 15).

Table 15. Stomach and tissue analysis could potentially inform indicators on the following MSFD Descriptors and Criteria.

Descriptor	Criteria
Descriptor 1: *Biological diversity*	1.1./1.3. Species Level 1.4./1.6. Habitat level 1.7. Ecosystem level
Descriptor 4: *Marine food webs*	4.1. Structure 4.2. Function
Descriptor 5: *Eutrophication*	5.1 levels 5.2 Direct effects 5.3 Indirect effects
Descriptor 10: *Marine litter*	10.1. Characteristics 10.2. Impacts

Ecosystem Approach

The need of implementing the Ecosystem Approach has been recognized worldwide during the last decades through a number of high level international agreements. But long discussions have occurred, first with the proper definition of the terminology used in that context and more recently with the difficulties in making it operational.

The Ecosystem Management *sensu stricto* is technically impossible, since there are intrinsic properties of the ecosystem and/or natural pressures affecting them which could never been managed. The main differences between the most commonly used terms, such as Ecosystem Based Management (EBM), Ecosystem Based Fisheries Management (EBFM) and Ecosystem Approach to Fisheries in terms of the sectors and the ecosystem components included, and also the evaluation processes and analytical tools to be used in each evaluation processes has already been explained in the literature (see Plagányi,2007), Link,2010) and Link and Browman,2014) for more details). All this information is summarized in

Table 16, modified from Link and Browman (2014), where their comparison with classical Fisheries Management (FM) has also been included.

Table 16. Tools and processes to make the EA operational (modified from Link and Browman, 2014), where IEAs: Integrated Ecosystem Assessments, ISAs: Integrated Stock Assessments; SA: Stock Assessments, WEMs: Whole Ecosystem Models; MRMs: Minimum Realistic Models; DSMs: Dynamic System Models; ESAMs: Extensions to Single-species Assessment Models; SAMs: Single-species Assessment Models.

	EBM	**EBFM**	**EAF**	**Classical FM**
Sector	All	Fisheries	Fisheries	Fisheries
Biological components	All	Community/ Whole ecosystem	Stock/Population	Stock/Population
Evaluation process	IEAs	IEAs (Fisheries)	ISAs	SAs
Analytical tools	WEMs	WEMs/MRMs/DSMs	MRMs/DSMs /ESAMs	SAMs

Different requirements in terms of stomach content analysis would then be needed depending on which level of complexity of the ecosystem approach is being implemented. For both, EBM and EBFM, the complexity of the tools used (whole ecosystem models such as Atlantis and EwE) in terms of the biological compartments of the food web requires a complete knowledge of the trophic relationships that exist between all the functional groups at different trophic levels. Moving towards the lower complexity levels of the ecosystem approach relaxes these needs, since these approaches are focused on one or few stock and generally multispecies models or extensions of single species models are used. This means that implementing EBM or EBFM would require collecting stomachs of individuals representative of all species or functional groups in the food web, whereas implementing EAF by using multispecies models or extensions of single-species models would require colleting only stomachs of individuals of the most important commercial species and their main predator/preys. The use of one or other approach will depend on the final objectives (tactical or strategical management; see FAO, 2008), but following the requirements of the European legislation in place, a combination of them would be required. Figure 8 shows different types of models that are used when moving toward the ecosystem approach and gives an idea of the level of complexity of each model in terms of the number of trophic levels included.

The previous paragraph implies that all the species or functional groups that coexist in a certain ecosystem should be sampled. However, this would be really hard to achieve for ecoregions with a high biological diversity such as the Bay of Biscay and Iberian waters. The use of the existing knowledge available in the literature about how these highly biodiverse systems work and which are the key species, not only in terms of their commercial relevance, but also in terms of their specific role within the ecosystem would be crucial for implementing appropriate sampling protocols. Higher sampling efforts would then be required in those regions where scarce information is available; but in addition to the sampling effort, the need of a proper computerization and compilation of the available information should be emphasized (see Huwer et al., 2014).

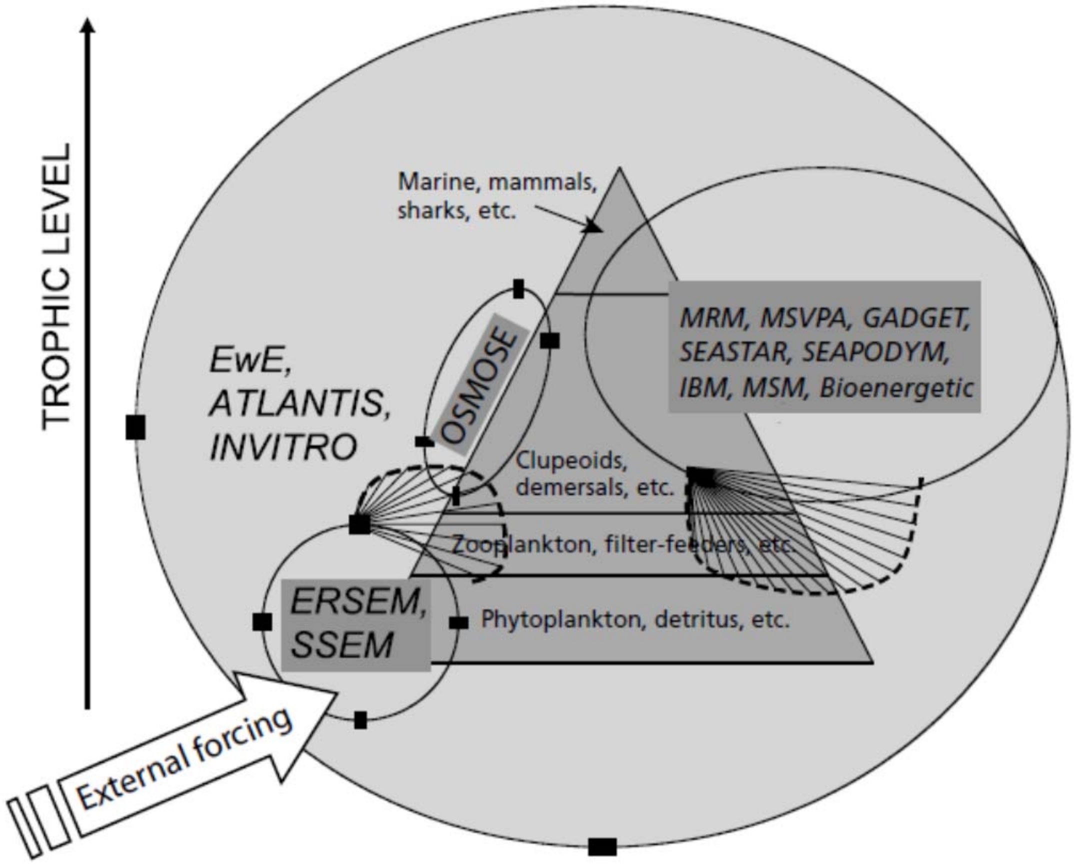


Figure 8. Schematic summary showing the trophic level focus of different multi-species models (Plagányi, 2007)

Stock assessment improvement

Fisheries management in European waters is based upon single-species plans that often have neglected biological interactions such as predator-prey relationships or the so-called technical interactions (e.g. by-catch, discards), and also any potential interaction with other environmental variables. Understanding the main predator-prey relationships is recommended if MS want to improve knowledge on the trophic interactions that affect yields and productivity. Fishing pressure exerted on a prey species might have a significant indirect impact on predator stocks or vice versa and incorporation of complex feeding behaviour into fisheries models is crucial if we are ever make realistic fishery predictions in the future and manage fish stocks on a sustainable basis.

At the very least, identification of key species and critical linkages or dependencies in commercial species will ensure the implementation of reliable multispecies models. The results of multispecies (and ecosystems) models can be adopted by managers as a strategic tool to assist in decision-making within a single-species management framework, also ensuring approaches that are compatible with an EAF.

From regional ecosystem models, historical data and studies and/or expert knowledge, key predator-prey relationships and trophic Interactions can be identified (e.g. key commercial piscivorous-forage interactions). Collecting data on the stomach contents to obtain a reliable estimate of quantitative dietary data (e.g. ration index) is crucial to improve the current stock assessment in European waters and for using multispecies models like deterministic Multispecies Virtual Population Analysis (MSVPA) or Stochastic Multispecies Model (SMS). Some studies also suggest that forage species assessed abundance has not always been reflected in predator´s diets.

This issue can be representative of prey selectivity or preferences by predators but can also be an indicator of the usefulness of stomach contents analysis to improve the assessment of forage species.

To improve estimates of natural mortality and MSY BRP´s of both commercial pelagic and demersal predator species in European waters it is also crucial to obtain more detailed information on diet description.

1.4 Type of data needed (end-users consultation)

Determining an appropriate sampling design is dependent on a clear understanding of the objectives for which stomach data is needed. As stated in the previous section there is a lot of information that can be collected from stomach analysis depending on which type of monitoring is required and on the complexity of the models that are being implemented.

Several end-users were directly consulted to assess the need for stomach contents data coverage using the project proposed questionnaire (Annex 1). The results from the questionnaires are summarized in Table 17. One of the respondents already has a demersal sampling program and a few institutions are involved in a regional coordinated single species program, but the majority do not have standardized diet studies. The need for a regular stomach sampling programme and data to feed ecosystem or multispecies models and MSFD directives was almost unanimous among the consulted end-users. It is worth noting that because the content of these questions was less clear, there was no answer that would specifically favour one ecosystem or multispecies model and which MSFD descriptors could be tackled. However, some end-users specifically mentioned the need for an up-to-date and regular parameterisation of existing multispecies assessment (e.g. cod). Moreover, a stomach sampling program is currently being implemented in one MS to answer to MSFD food web descriptor and D10 (marine litter) is also mentioned as easily monitored with no, or very little, additional costs for a sampling programme.

Regular stomach sampling is needed in a range of species groups, from pelagic keystone species to migratory apex predators or from highly abundant species to declining stocks. As some users stated that they try not to focus on single species of immediate interest but on all species of major ecological interest, others define priority species according to their importance in terms of biomass or abundance. Overall, the commercial fish were given the most importance and among these, mid-level to highly influential top predators (e.g. cod, whiting, hake). Small pelagic species are also mentioned as priority species, but the results from egg and larval predation studies are more difficult to produce firm conclusions. Additionally, it is our impression that the preferred sampling design is dependent on the number of “interesting” species, ranging from intensive species specific, an all inclusive or a rollup sampling scheme.

Most end-users had difficulties in estimating the sampling cost (including collection, analysis, data logging, consumables and labour) and the results ranged from 7€ to 880€ per individual stomach. Estimated cost is considered highly dependent on the species considered and the type of diet analysis. Much of the information is collected in opportunistic diet studies within the framework of ecosystem surveys, environmental impact assessments, observer programmes, ecology and postgraduate studies.

It is also noticeable that all responding institutions have stomach data mostly from commercial pelagic and demersal species and sample stomachs occasionally during individual studies. This considerable body of historic data could be integrated in a common regional stomach contents database (“most benefits for least cost”) to feed existing models and improve understanding of long-term community interactions within each ecosystem.

Besides the valuable input to ecosystem/multispecies models, diet information was considered relevant to many ecological studies, including essential fish habitats (EFH), physical-biological interactions, growth rate studies, and long term changes in food web structure, etc.

Table 17. Stomach sampling coverage and data needs from individual questionnaires (available in Annex 1). *Per stomach, including collection, analysis, data logging, consumables, labour; [1] Stomach sampling program will be implemented under MSFD D4; [2] demersal predators data from 2010-2012 available in database.

Country	Institution	Active sampling program		Available historical data			User needs						Cost (€)*
		Species / Area	Year	Species (nº) / Area	Year	Database	Sampling program	Species / Area	Multi species models	Ecosystem models	MSFD	Other	
Spain	IEO	Demersal species / Div. VIIIc	1990 - present	Demersal / Div. VIIIc	1990 – present	Yes	No	–	No	Yes	Yes	PhD Thesis and other scientific publication	NA
France	IFREMER	No[1]	–	Demersal and pelagic (18) / Div. VIa, VIId	2009;2013	Yes	Yes	Demersal and pelagic / French coast	Yes	Yes	Yes	Trophic ecology research	180
Portugal	IPMA	No	–	Demersal and pelagic (29)/ Portugal EEZ(Div. IXa, Azores and Madeira)	Scattered (from 1987–2014)	No[2]	Yes	Demersal and pelagic / Div. IXa	No	Yes	Yes	Food web structure/ ecology PhD´s	NA
Germany	TI-SF/OF; UHAM	Cod / Baltic SD 22, 24, 25	2013–present	Demersal (20) / North Sea and Baltic	1992, 1996,1997, 2001, 2003–2006, 2013	Yes	Yes	Commercial species / North Sea and Baltic	Yes	Yes	Depend final indicators (D4 and D10)	Feeding ecology	15
Poland	NMFRI	Cod / Poland EEZ	2015–present	Cod,flounder / Poland EEZ	2006–2014	Yes	No	–	Yes (Cod)	-	-	-	7
UK	CEFAS	No	–	DAPSTOM	1837 – 2012	Yes	Yes	Commercial species / North Sea, Celtic Sea and Irish Sea	Yes	Yes	Yes	Long-term change in food webs	NA
Sweden	SLU	No	–	Demersal (6) / Baltic Sea, Skagerrak-Kattegat	1980, 1990, 2012–2014	Yes	Yes	Cod / Baltic Sea	Yes (cod; sprat;herring)	Yes	Yes	Growth rates and condition of fish (Cod)	NA

Italy	OGS	No	–	Pelagic (5) / Adriatic	2002–2003, 2006–2014	No	Yes	Pelagic keystone species, demersal, tunas and large pelagic predators	NA	NA	NA	NA	NA
Belgium	ILVO	No	-	Demersal, flatfish (4) / North Sea	2009–2014	Yes	No	-	No	maybe	maybe	Impacts of trawling and wind parks	190
USA	NOAA (Southwest coast)	tunas/highly migratory large fish	1997–present	tunas/highly migratory large fish	1997–present	Yes	No	-	Maybe	Yes	Not applicable	Establish EFH	880
USA	NOAA (Northeast coast)	demersal, pelagic, elasmobranchs	1973–present	demersal, pelagic finfish and some elasmobranchs (>150)	1973–present	Yes	No	-	No	No	Not applicable	Diet-climate studies, BRP´s	NA

1.5 Data collection and sampling design

1.5.1 Review of data collection methodologies and data sources

Surveys and at sea sampling programmes

Historically, different research surveys have been used to collect stomach content data but it has never been the main objective of these research surveys. Again, specific sampling programmes and protocols to collect these data will need to be modified and/or extended to provide surveys that cover all the needs of relevant end users. As the collection of stomach contents data is expensive if dedicated surveys are required, it is recommended that samples should be taken opportunistically on existing DCF trawl surveys. A common sampling method and gear is ideal in providing a standard index of diet over time and across regions that would facilitate inter regional comparisons. The well stablished International Bottom Trawl Survey (IBTS), Beam Trawl Surveys (BTS) and pelagic surveys are presently widespread in European waters and have been consistently used to assess abundance of the main species. With the application of a common sampling protocol, the existing research surveys can be the current standard for a comparable and comprehensive diet description (see section 10 for more details).

Researchers must be aware that samples of fish collected from a given region at one time may not provide diet information typical for that species at other times and in other locations. However, the rigorous collection of diet data throughout the existing annual research surveys function adequately as an annual index for statistical analysis and multispecies models at the specific region.

Other sampling gears in commercial vessels may also be appropriate for collecting stomach samples, but their uniform use is generally not common enough to permit spatial (regional), temporal (annual) comparisons. Stomach sampling on board fishing vessels involves a considerable amount of work and it would greatly rely on the way the observer programmes operate and dependent on the presence of appropriate staff members between coordinators and observers; an additional observer on board would probably be required in each national observer's programmes, with a significant cost increase. There would probably be a constraint on the number of analysed stomachs and samples would mostly be from fish catchable sizes not ensuring a random selection within size classes. Additionally, the geographical coverage of fishing vessels is in most cases not reflective of the spatial distribution of the predator species.

However, if it is considered necessary to assess diel variations, predator's seasonal behaviour throughout the year, as well as obtaining samples from nearshore areas or areas that cannot be trawled, commercial on board sampling in ships of opportunity offers some advantages by filling spatial-temporal gaps. Differences in gear selectivity (time of day, depth) should be taken into account when combining the data for diet description and active methods such as seining and trawlering are preferable. Impoundments or static gears can also be used as long as the net sets are 24 hours or preferably less in duration because of the digestion that occurs while fish are held. Besides the specific objectives, if such sampling is required, the sampling protocol should follow the same standards of the research survey protocols (section 2.3) and results from different gears should only be combined if data show similar properties (e.g. percentage of empty stomachs is different between baited non-baited gears).

Visual and genetic analysis

Stomach contents analyses provide a valuable source of information on predator–prey relationships. However, identification of prey items in stomach contents can be difficult , and frequently result in short-term dietary snapshots caused by short digestion times of soft-bodies and small organisms. High quality analyses of stomach contents are an essential part of food web understanding that cannot be duplicated by other methods.

The accurate depiction of the diet, with the identification of the prey taxa in dietary samples, however, can be time-consuming and unsatisfactory using conventional morphological analysis (e.g. Pompanon et al. 2012). Several methodologies were tested to validate and apply DNA-based approaches to describe fish diets across the scientific community. Results showed that molecular techniques were more precise than traditional visual inspection and could provide insight into diet preferences for even highly digested prey that have lost all the morphological diagnosing characteristics, identifying prey taxa after much longer digestion times than possible with visual methods (up to 24 hours). Despite being high-priced systems some studies suggest new techniques that provide efficient, easy to use and cost-effective tools for assessing the trophic ecology of piscivors (Martinez et al, 2011; Thalinger et al., 2015).

Metabarcoding of prey items in predator guts provides a valuable tool for characterizing previously unknown trophic interactions. The method relies on the availability of a DNA reference library to which the prey sequences can be matched. There are some sequence databases available to enable the identification of a large proportion of digested fish, crustacean and molluscs, namely for Cytochrome c oxidase I (COI) barcode region (e.g. The Barcode of Life, http://www.boldsystems.org; Fish Barcode library, http://www.fishbol.org; Moorea BIOCODE project, http://www.mooreabiocode.org). As much of the digested prey remains unidentifiable, the DNA barcoding identification will become increasingly useful as more vertebrate and invertebrate reference sequences are added to the libraries. Rapid developments in molecular biology are providing high-throughput sequencing techniques that enable DNA identification of numerous taxa in large numbers of gut samples (Dunn et al., 2009; Leray et al., 2012). Also, the molecular methods have often been used to give presence/absence information, but the metabarcoding can be applied semi-quantitatively to estimate relative abundances within a sample (e.g. Amend et al. 2010; Murray et al. 2011).

Overall, the genetic approach and DNA barcoding is no doubt a useful tool and should be considered as a random method for identifying predator's digested, rare or small prey species, detecting the presence of eggs and larvae and/or for a rapid screening of a large number of predators. The major constraint of these techniques is the high unit cost, which may decline as the new technologies become mainstream.

Biochemical methods

Biochemical tracer techniques (stable isotope and fatty acid analysis) provide well established methods for investigating trophic relationships that offer several advantages over stomach content analysis. Stomach content analysis are a real representation of the diet of a studied species and are the basis of the knowledge of predator-prey relationships. On the contrary, isotopic signatures (usually carbon and nitrogen) of predators integrate diet variations in time. These two methods are complementary and inseparable to study the diets of species and to estimate the most precisely possible trophic level of species

These techniques focus on assimilated versus ingested prey material, and can be used to identify food web sources, which cannot be achieved through stomach content analysis. Depending on the tissue type and the biochemical compound analysed, dietary information is integrated over periods of days to years. Trophic information can be acquired from preserved soft and hard tissues and thus enable retrospective analyses using archived material (Young et al., 2015).

These techniques have already been performed in occasional studies in some MS and have been suggested as of crucial importance to improve food web characterization and tackle the (Link, 2010) challenges of the MSFD. Moreover, accurate trophic level data could support the use of multispecific and/or ecosystem models.

1.5.2 Coverage of the current sampling

Demersal community coverage

Several ICES coordinated annual surveys could, with some additional effort, be used to perform an opportunistic stomach sampling programme as diet description should afford compatibility with the temporal and spatial standards of research surveys . The International Bottom Trawl Survey (IBTS), the Baltic International Trawl Survey (BITS), the Mediterranean International Trawl Survey (MEDITS) and Beam Trawl Surveys (BTS) already have well established standards that can provide species-abundance distribution at adequate spatial scales. ICES statistical rectangles as used by some of these surveys seem to be adequate to assess the observed spatial differences in consumption due to the generally patchy distribution of predators and their prey. This proposed stratification can differentiate habitats where diet has been shown to differ with water depth or distance from shore. However, spatial divisions can be regionally aggregated if there is no significant difference in feeding behaviour.

IBTS surveys, covering most of the North East Atlantic is show on Figure 9 and the MEDITS programme provides trawl standardized surveys in the Mediterranean Sea (Figure 10).

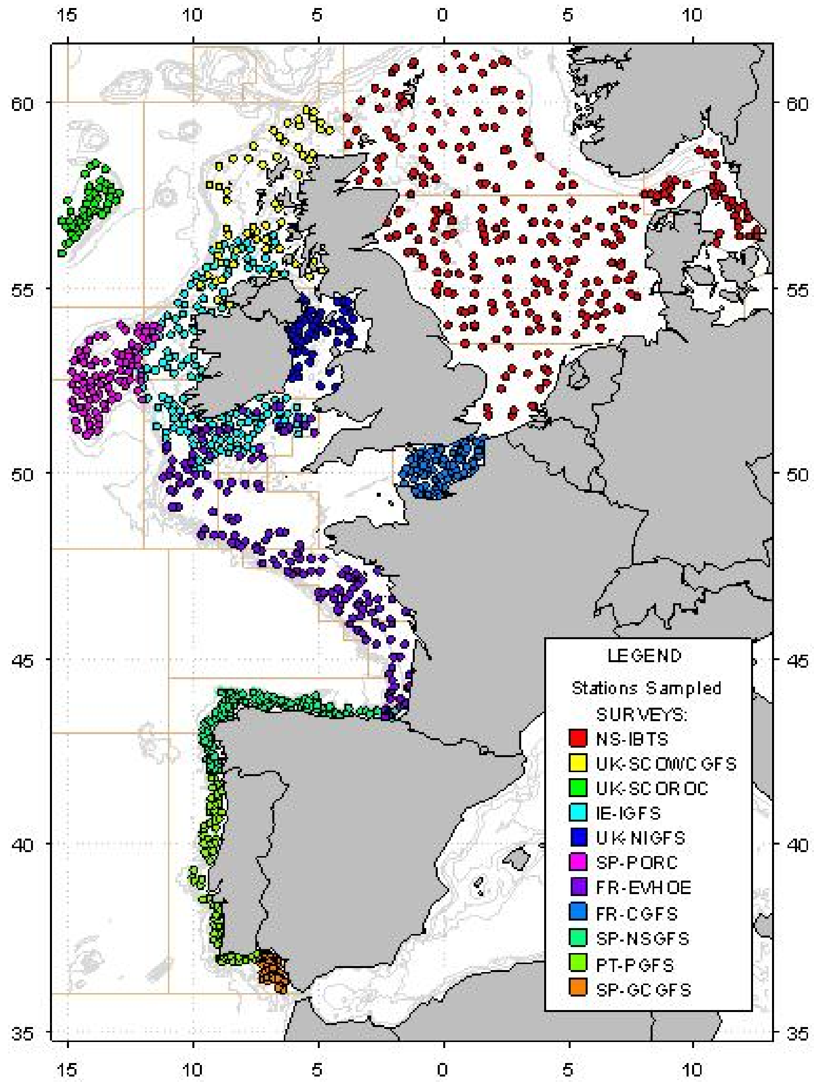


Figure 9. IBTS surveys carried out in the North Eastern Atlantic and North Sea areas.

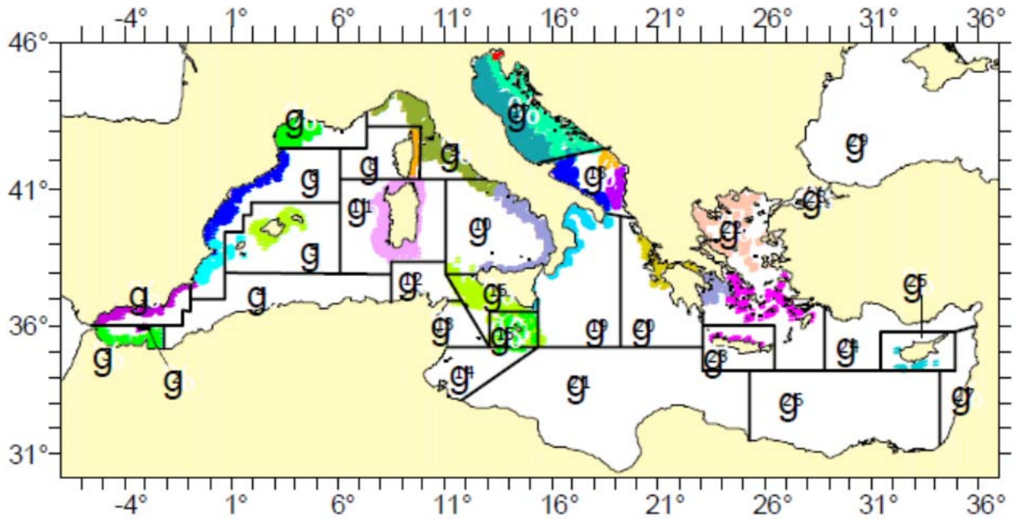


Figure 10. MEDITS survey coverage in the Mediterranean Sea. The limits and numbers are those of the GFCM-GSAs. The coloured areas are those of the main strata and hauls.

Pelagic community coverage

Pelagic surveys are not that coordinated as demersal surveys, though in total they cover a similar area. Different methodologies are used: some surveys use acoustic techniques to estimated abundances of pelagic species and some others sampling eggs/icthyoplankton and apply the Daily Egg Production Method (DEPM) afterwards. A complete list and description of this type of survey can be found in the STECF report (SGRN 2010). Both acoustic and egg production surveys might be good candidates for a potential stomach sampling programme, since they both have associated samplings targeting fish (adult or juveniles depending on the survey). Using both surveys would additionally provide a better temporal resolution of the information, which for some species could be representative of different life stages. Investing some effort in harmonization of these types of surveys is recommended.

One of the biggest limitations of sampling stomachs of pelagic species could be the planktivorous nature of some of these species, especially the small pelagics which mainly feed on phyto- and zoo-plankton species. This leads into a more time consuming visual analysis, and so the cost of the sampling will also be increased. A potential solution for this could be the use of genetic identification methods, but they should be calibrated some years in advance. It has also been suggested that, in addition to the more challenging sampling, the results from egg and larval predation studies are also more difficult to interpret.

1.6 Data collection

1.6.1 Existing stomach databases in Europe

Stomach sampling in European waters has been approximately annual in some areas, while in other areas (e.g. the Year of the Stomach in North Sea) a large effort has been made sporadically. As such, a number of coordinated fish stomach databases do exist in Europe but the majority encompass a limited selection of species or cover a very discrete period of time. The more complete and widespread dataset is the DAPSTOM database (https://www.cefas.co.uk/cefas-data-hub/fish-stomach-records) which represents one of the largest and most diverse compilations of marine foodweb data anywhere in the world with 226.407 records derived from 449 distinct research cruises and several research projects spanning the period 1837-2012 and several ICES sub-areas. There are several reports available detailing the contents of the database (e.g. Pinnegar, 2014a, Pinnegar, 2014b). The ICES "Year of the Stomach" database for the North Sea (http://www.ices.dk/marine-data/data-portals/Pages/Fish-stomach.aspx) is another extensive dataset with reported 1.149.608 records, primarily based on stomachs collected during sampling campaigns in 1981 that covered a handful of species in the North Sea and a follow up data collection performed in 1985-1986, and a decade later in 1991. This ICES dataset provides information on 35 species, although detailed data is only available for nine. A similar coordinated ICES dataset exists for cod in the Baltic Sea that has been documented in ICES (1997). In the Barents Sea, a combined database exists for Norway and Russia (Dolgov et al., 2007), but large areas of the north-east Atlantic still remain neglected.

1.6.2 Other Databases

Several integrated stomach contents database are available worldwide, including:

- Food web Dynamics Programme (NOAA) – Integrated stomach database for the US southwest and northeast coast. Stomachs of major species such as Atlantic cod, haddock, silver hake, yellowtail flounder, winter flounder, Atlantic herring, and Atlantic mackerel are individually preserved, stomach contents of all other species are examined and identified on-board. The programme requires that a specific number of fish are sampled per length class (specified by species) and major fish species are sampled at every station. Shipboard processing includes volumetric measurement of prey, percent composition, prey number, and prey lengths.

- FISHBASE (http://www.fishbase.org/search.php) - Global information system with extensive information on more than 32.000 species and subspecies of fish. Compiles the information of several stomach contents studies and trophic level estimations.

- The California Current Predator Diet Database (CCPDD) - Synthesizes data from published records of predator food habits over the past century. The database includes diet information for 100+ upper-trophic level predator species, based on over 200 published citations from the California Current region of the Pacific Ocean, ranging from Baja, Mexico to Vancouver Island, Canada (Szoboszlai et al., 2014).

- PANGEA open-access data portal (www.pangaea.de) - Includes several types of stomach information as for example compilation of stomach content data for pelagic fish species (herring, blue whiting, mackerel, albacore and bluefin tuna) in the northeast Atlantic.

1.6.3 Data analysis and quality evaluation

Data quality evaluation of fisheries data encompasses the statistical soundness of the sampling design, the outcomes of implementing the scheme, how the data are managed, and how the data are analysed. In a practical approach, quality evaluation can be distilled down to two core elements:

a) An evaluation of whether the sampling program is designed and implemented, and the data managed and processed, in a way that follows agreed sets of standards. A programme meeting these standards is in principle capable of providing the desired standard for data quality.
b) An evaluation of the quality of the data that have been collected, using diagnostics and quality indicators that identify potential (or known) bias, and those that provide estimates or indexes of achieved precision.

During recent years many workshops have been carried out under the umbrella of ICES (WKACCU, WKPRECISE, WKPICS) covering aspects related to data analysis and quality evaluation. Best practices guidelines, protocols and quality check standards have been provided. Under the fishPi project and specific work package (WP4) focused on these issues too. All the output from this work is applicable to stomach content monitoring programmes and it is recommended that these reports are used as reference with respect to these issues.

2 Eco-region guidelines for a protocol of stomach data collection

Defining a sampling plan using a single stock as case study has no added-value as it does not provide enough information to show accurately how MS could coordinate and adapt a multi-species regional sampling programme to support the move towards the EAF with sufficient, statistically-sound samples of all the key species needed for food web characterization and for quantifying predator-prey relationships and predation rates within different types of multispecies and ecosystem models.

From the previous sections there is already an idea of what information can realistically be collected that meets the data needs of the scientific community. However, the development of stomach sampling, the availability of data and level of the local trophic dynamics knowledge is very different between MS (namely northern countries and southern countries) constraining the development of a common stomach sampling protocol. Nevertheless, this protocol tries to ensure that the composition of predator diets are reported in standard units and quantified to provide results comparable across regions and over time and should easily input into a multiplicity of ecosystem/multispecies models. In addition to diet data, stomach contents analysis can also be useful in many thematic areas relevant to the marine environment, that are not usually covered by some of the existing monitoring programs tools. To optimize the scientific information provided by the collection of stomachs, some general recommendations and sources are also given to provide guidance if additional monitoring is required on, for example, biotoxins contaminants and plastics monitoring under the MSFD.

Roughly these guidelines for a stomach sampling protocol are separated into topics: 1) provide guidelines for MS to consider priority stocks for stomach sampling and identify areas and stocks for future MS sampling coordination at a regional scale; 2) assess the development and availability of stomach sampling studies between MS, and; 3) establish a stomach collection protocol that will ensure data collection is prioritized and harmonized (e.g. considering established databases, namely DATRAS and the ICES "Year of the Stomach" database).

2.1 Consider priority stocks and data needs to conduct stomach content analyses to support our knowledge of potentially important predators for which the diet is presently poorly known

Determining the species to be sampled is dependent on the specific user needs and a clear understanding of the objectives of the diet study. Based on the regional expertise, MS should prioritize predator species for stomach sampling taking into account several criteria:

i) Is the species sampled for maturity and age in existing surveys (less effort for an additional stomach sampling)?
ii) Has the species only been subject to very limited stomach analyses?
iii) Is there possible significant change in diet since the last sampling study in the region?
iv) Is there a potentially large impact on key and commercial prey fish (recognisable top-down control from ecosystem models or ecology studies)?
v) Will sampling deliver a recognisable increase in knowledge of the ecosystem?

2.2 Bay of Biscay and Iberian Waters Case Study – France, Spain and Portugal (Recommendation at Regional scale)

The Bay of Biscay and Iberian waters is an appropriate eco-region (Figure 1) to tackle stomach data collection integrated across three MS. The development of stomach sampling and data availability is very different between these MS, providing a good example on how the different collection protocols can be applied (Table 18).

Table 18. Current coverage and use of stomach data in the Bay of Biscay and Iberian waters by MS (details in text).

MS	Historical data	Existing Database	Common exchange format	Sampling program	Stomach data used in:	
					MSFD	Ecomodels
Spain	Yes	Yes	No	Yes	Yes	Yes[2]
France	Yes	Yes	No	No[1]	Yes	Yes
Portugal	Yes	No	No	No	Yes	Yes[2]

[1] sampling program under development to answer MSFD D4
[2] sub-regional model

Spain (IEO & AZTI)
A stomach data sampling program developed by IEO started in 1988 and continues during the annual demersal survey on board the RV "Cornide de Saavedra". The survey covers the Cantabrian Sea and Galician waters. The survey is performed every autumn and stomachs are analysed quantitatively on board using a trophometer to measure the volume of the stomach content. Prey species in the stomachs are determined to the lowest possible taxonomic level. Decapod crustaceans and molluscs are measured individually while other invertebrates are counted and maximum and minimum sizes are recorded (if possible). Hence, approximately 10,000 stomachs are analysed every year for the main fish predators. These data were used to define trophic processes in GADGET and EwE models for Atlantic areas around the Iberian Peninsula. Hake stomachs have been visually analysed at AZTI since 2009.. Results from a more recent pilot study funded by the Basque Government, where metabarcoding techniques have been used to carry out stomach content analysis, are already available at AZTI.

Portugal (IPMA)
Irregular stomach samplings were made in recent years during the annual demersal survey on board the RV "Noruega” on the main predator species. Depending on the available time/personnel, the stomachs were analysed quantitatively on board or the whole fish was collected for later analysis. The extent to which prey had been digested was recorded and species determined to the lowest possible taxonomic level and, whenever possible, measured and weighed individually. Different studies were carried out throughout the years often with different goals and rarely have been carried out over a long period of time in a consistent manner. There are also concurrent diet studies among institutions and the reporting format is not standardised. IPMA does not have an integrated stomach contents dataset. However, there are a lot of published and unpublished data (e.g. grey literature, thesis) on stomach sampling from a diversity of species that can be analysed as to its

usefulness and standardised into a dataset. Some of these historical data have already been reviewed for the MSFD D4 initial assessment and regional EAF studies (e.g. Cardoso et al., 2015).

France (IFREMER)
There is not a coordinated stomach sampling program in France yet, but the protocol regarding Descriptor 4 on food webs is about to be approved under the MSFD program framework. It includes a list of species to be sampled in each of the French marine ecosystems (the English Channel, the Atlantic coast, and the Mediterranean Sea). Once the list of species is confirmed by the ministry, France will have to implement a sampling program to analyse the stomach contents of these species (mostly defined according to their importance in terms of biomass or abundance). Information obtained from occasional studies is available in some of the laboratories' databases, providing information about the length, weight, age, sex and maturity of some selected predator species, and the number and weight of all identified prey. Visual analyses are used for most of the studies, but more recently biochemical methods have been used in some cases. All this information is already been used at both, regional and national level for the implementation of multispecies and ecosystem models, but also to provide indicators for the MSFD.

Section 2.1 already introduced some guidelines to consider priority stocks and data needs to conduct a stomach content study. As a rule, species should be selected according to specific user needs. Ideally, the one most "important" species would be selected for a practical and cost-effective regional sampling program (similar to cod in the Baltic and North Sea), as this is not informative enough for this high biodiversity region we propose that, dependant on the user needs, the priority predator species should be chosen according to:

i) Most of stock distribution overlapping the case study region MS and with available estimates of SSB and F
ii) Common species sampled for maturity and age in existing surveys or already sampled in an existing stomach programme
iii) Relevant species to monitor under MSFD D1, D8, D10
iv) Large impact on the ecosystem (e.g. biomass level; key species; apex predators)

Table 19 provides some examples of stocks that could be considered in a regional sampling programme for the NE Atlantic.

Table 19. Examples of NE Atlantic stocks and selection criteria to be considered in a regional sampling programme.

ICES Stock	Area	MS	SSB and F/ BRP´s	Common priority species[1] /survey	MSFD Descriptor	Ecosystem comments	Considered for sampling programme
Southern Hake	VIIIc + IXa	PT, SP	Yes / Yes	Yes / IBTS	D1, D3, D4	large biomass, recovery plan,top food web,generalist; OSMOSE	Yes
Northern Hake		SP, FR	Yes / Yes	Yes / IBTS	?	?	?
Sardine	VIIIc + IXa	PT, SP	Yes / No	No / IBTS	D1, D3, D4, D10 biotoxins	present low biomass,key species; ECOPATH	maybe (planktivore; challenging sampling)
Southern Horse Mackerel	IXa	PT, SP	Yes / Yes[2]	Yes / IBTS Pelagic acoustic surveys, MPDH	D1,D3,D4	present high biomass, preferential preys	Yes
Horse Mackerel	Subarea VIII Divisions IIa, IVa, Vb, VIa, and VIIa–c, e–k	SP, FR, others	Yes / Yes	Yes / IBTS acoustic surveys, MPDH	?	?	?
Anglerfish	VIIIc + IXa	PT, SP	Yes / No	Yes / IBTS	D1, D3, D4, D8	Top food web, generalist	maybe
Blue whiting	Subareas I–IX, XII, andXIV	PT, SP, FR	Yes / Yes	Yes / IBTS	D1,D3,D4	high biomass, key forage specie	Yes
Mackerel	Subareas I–VII and XIV and Divisions VIIIa–e and IXa	PT, SP, FR, others	Yes / Yes	Yes / IBTS acoustic surveys, MPDH…	?	?	?
Anchovy GOC	IXa	PT,SP	Yes / Yes	Yes / acoustic surveys, MPDH…	D1,D3,D4, D10? biotoxins	Key commercial species	maybe (planktivore; challenging sampling))
Anchovy BoB	Subarea VIII	SP,FR	Yes / Yes	Yes / acoustic surveys, MPDH…	D1,D3,D4, D10? biotoxins	Key commercial species	maybe (planktivore; challenging sampling))
Northern Sardine	Divisions VIIIa,b,d and Subarea VII	SP,FR	No / No	Yes / acoustic surveys, MPDH…	? biotoxins	?	No
Chub mackerel	No assessment	PT, SP, FR,	No / No	Yes / acoustic surveys, MPDH	D1, D3, D4;	High biomass, northward migration/pelagic competitor	No

[1]species sampled for age and maturity
[2]BRP´s proposed in 2016

2.3 Stomach sampling protocol

2.3.1 Include all appropriate MS historical stomach content information into a common database

As part of the contract agreement for EU 'Lot' project MARE/2012/02 – SI2.632887, it was recognised that "substantial historic stomach data already exist in the partnering institutes, much of which could be used in the estimation of multispecies interactions with much lower cost than that incurred by sampling new stomachs (Pinnegar, 2014b). The Baltic and North Sea areas, already made a large effort to retrieve such data and make them available to multispecies models in the format used by ICES in their "Year of the stomach" database.

The assessment, collection and transformation of existing historic stomach data into a dataset with the appropriate format will provide insight on predators for which the diet is presently poorly known. Historical data would also provide low cost stomach samples to be used in the investigation of the temporal stability of predator preferences and a significant improvement for the existing stomach information in European waters.

2.3.2 Stratifying by Region

Stomach samples should be collected opportunistically on existing research surveys as specified in section 1.5.2 using the established spatial stratification.

2.3.3 Stratifying by Season and Hour

Stomach sampling should be carried out at hauls of opportunity in existing research surveys (e.g. autumn IBTS)

Although, several predator-prey interactions take place during night time, most of the fishing stations are only carried out in daylight on standardized scientific surveys, as catch rates can be very different in the dark under certain conditions.

Stomachs samples ideally should be collected in hauls over distinct daytime periods (e.g. morning, lunch, afternoon)

There is evidence of diel and seasonal dependent diet composition, however, opportunistic sampling throughout the existing annual research surveys function adequately as a comparable annual diet index.

2.3.4 Stratifying by Species

In addition to environmental factors, prey availability, habitat use, and differences in diet have been attributed to predator species, age, size and life stage. To ensure a random selection across the latter attributes, stomachs should be sampled individually for each predator using fish selected for maturity and age sampling whenever possible.

2.3.5 Stratifying by sample frequency

As stated previously, several key species should be selected for regional stomach sampling. To avoid excessive effort and, most probably, an unrealistic sampling strategy for several MS, it is much more cost-effective to use a low frequency sampling regime across the fishing stations. Intra haul correlations among prey species, points to little gain in sampling a large number of stomachs at a station (ICES 2006, 2007). To account for spatial and temporal variability it, is much more efficient and cost-effective to sample only a few fish at each station and increase the number of stations.

Sample frequency is dependent on user needs and region specific (Northward systems: low biodiversity and/or Southward systems: high biodiversity)

a. Index sampling by collecting samples on an annual basis for a few predators and rotate each year can be viable to provide all inclusive stomach data. Sampling two to three species each year.

b. Another option is to have a sampling regime whereby each species is sampled each year but at very low sample sizes. This method used is efficient to perform in opportunistic sampling during existing surveys (WGSAM 2010).

2.3.6 Stratifying by length class and sample size

Determining the adequate number of stomachs to precisely describe the diet of a species is very difficult and despite its importance there is very little literature on the subject as this issue is rarely mentioned in diet studies. In general, species should be sampled for equal length classes spanning the historic observed length range to account for ontogenic differences in the diet. To avoid excessive sampling effort, the number of stomach samples should be dependent on the number of individuals that are selected for maturity and age sampling in each selected species.

Determining when enough samples have been collected to properly describe the diet or to compare diets is constrained by *a priori* knowledge of *cumulative prey curves*. If available, the use of cumulative prey curves is recommended; as sample size increases, variation in prey species richness decreases, reaching an asymptote. However, cumulative prey curves should not replace the need to assess sample size sufficiency for performing diet comparisons.

Generalist and specialist species have different sample frequency needs to properly assess diet information.

Laboratory processing of large numbers of samples has sometimes proven to be too costly. To avoid an excessive effort and, most probably, an unattainable sampling strategy, a parsimonious sample size can be established from approximately 10 individual/size class. within the fish selected for maturity and age in each haul.

When defining equal length classes be aware that larger /long lived/slow growing species should have bigger size intervals than short lived/fast growing species.

Because growth can vary regionally and most assessment models are age-structured it is important that the average and range of ages are represented in each size group on a regional level. A length frequency analysis is recommended for determining appropriate size classes that accounts for each species length-age curve.

The establishment of length classes can also be dependent on species specific observed ontogenic changes.

The time available for the collection and analysis of stomach samples on-board is often limited as no additional staff are allocated for the collection of stomachs in the existing surveys. Stomachs can be removed and quickly frozen individually in plastic bags with an appropriate label. If needed and to reduce the extra work, samples can be gathered by collecting whole fish instead of taking out the stomachs. The collection of stomach samples also allows diet contents to be used to monitor other parameters (see section 2.3.8).

To provide a quantifiable measure of feeding fish among these samples, care must be taken to obtain reliable data on (adapted from ICES, 2010):

i) Everted stomachs. Some fish have everted stomachs. Since it not known whether these stomachs contained food or not, such fish must **not** be used for stomach samples

ii) Regurgitated stomachs. Some fish have regurgitated all or part of their stomach contents and these fish must not be collected for analyses. However, the number of regurgitated stomachs encountered during the examination must be recorded to ensure that the proportion of feeding fish in the sample is accurately defined. In practice, it is often difficult to tell whether regurgitation has taken place, but in situations where the stomach is flaccid or distended, but contains little food, the size of the gall bladder is a useful practical indicator of the recent feeding history of the fish. When collecting the stomach sample, the state of the gall bladder should be recorded to help differentiate between empty and regurgitated stomachs. Shrunken or elongated pale green bladders points to recent feeding and round or elongated dark green bladders indicates non feeding fish.

iii) Stomachs of feeding fish showing no signs of regurgitation. These should be collected for analyses. It should be noted that not all feeding fish have grossly distended stomachs, i.e. feeding does not necessarily mean full.

iv) Empty stomachs.

v) Stomachs with only indigestible skeletal remains (e.g. polychaete bristles, mollusc shells and opercula, fish bones and otoliths etc.).

The material collected at sea to meet the sampling targets should originate in feeding fish showing no evidence of regurgitation and from non-feeding fish. Sampling should continue until the established number of samples by length class is achieved.

Regurgitated or empty stomachs do not provide information on the list and prey consumed but they have to be recorded to compute an index of ration providing a quantifiable measure to be used in several multispecies models.

Some fish/invertebrates prey species identification is assumed to be more accurate when stomach contents are fresh.

2.3.7 Laboratory stomach contents analysis

There are three basic steps to analyse stomach contents:

i) Weigh the entire stomach content.

ii) Identify fish and invertebrate prey to the lowest taxonomic level when possible.

iii) Digestive status is recorded and length of fish and invertebrate prey are measured or estimated to nearest cm (rounded down). The weight and/or volume of individual prey or prey groups are recorded.

Percentage composition by numbers, weight and/or volume are frequently used measures to evaluate the importance of different prey types in diet studies. It is recommended that the information collected should be sufficient to use more than one method to describe diet

Stomach sampling has to be performed by very experienced personnel. Each institute has to be able to handle this work with proper formation.

Several methods can be used to describe stomach contents including occurrence, numerical, volumetric, and gravimetric indices, each method being subject to varying degrees of bias and accuracy. There has been some debate over how to estimate precision and best represent the composition in fish diet studies. Numerical methods usually provide more robust measures with considerable economic and logistic benefits but gravimetric/volumetric indices, despite more time consuming, provide more detailed information for the estimation of the total amount of food consumed by a population.

Quality control regarding the identification of prey species between MS (e.g. photo analysis and comparison).

Standardize digestive status (ranging from 0 "pristine" to 2 "very digested) between MS of individual prey items (e.g. photo analysis and comparison to standardize digestive status assessment).

2.3.8 Stomach studies to answer other user needs

Trophic level (stable isotopic analysis)

Detailed analysis on the trophic level of species according to size (ontogenetic shifts) are missing in some European areas. Complementary to stomach contents it is suggested that tissue samples should be collected from the most abundant species within research surveys. White muscle samples should be collected from the dorsal musculature for each equal length class spanning the species observed length range, placed in a vial and immediately frozen. These techniques are expensive and to avoid excessive effort and potentially unattainable sampling strategies an irregular low intensity sampling regime can be implemented whereby one to two species are sampled each year. Stable isotopic analysis can also be performed from dried preserved tissue samples across a 10 year window.

MSFD

There are also possibilities for synergies between diet studies and the monitoring of human pressures and impacts. Fish stomachs, provide an accurate indication of plastic levels in the marine environment and have, as such, been named as an indicator for GES under MSFD Descriptor 10. The presence of micro- and nanoplastics in fish stomachs is an indicator of marine litter. A preliminary protocol for monitoring of plastics in fish stomachs in the marine environment was made available by the ICES Working Group on Pathology and Diseases of Marine Organisms (WGPDMO). There has been no such monitoring of plastics previously, so this protocol will be reviewed at regular intervals and improved on the basis of experience data (ICES, 2015a). The implementation of this protocol requires that the selected stomachs are collected on-board for laboratory analysis.

Biotoxins

Subsamples can be collected from the fresh stomach contents and tissues sampled from impacted species (pelagic) and frozen for laboratory biotoxin analysis.

2.4 Cost of stomach collection and analysis

As already mentioned throughout this report, the cost of a stomach sampling programme is highly dependent on the species considered (e.g. pelagic diet description more challenging), type of diet analysis (e.g. MSFD; contaminants) and intensity of the sampling scheme (all inclusive, rollup scheme). A cost estimate of sampling stomachs will also have high uncertainy given the scientific research vessel, variable use of volunteers and paid scientists, the amount of time required per stomach, existence of experienced personnel within each institution, etc.

2.5 Examples of cost effective sampling

The most effective sampling scheme and intensity is dependent on the specific user needs. Many of the guidelines presented here benefit from opportunistic sampling in internationally coordinated survey programmes inside the DCF and add on fish diet sampling to minimize direct costs. The possibilities for synergies between a stomach sampling programme and the monitoring of human pressures and impacts under the MSFD, and surveillance of marine biotoxins should also be strongly considered. This valuable additional information could be analysed at very little additional cost providing data relevant in many areas of scientific knowledge and with significant added value for the fishing industry, economies and human health.

2.6 Formats

After stomach contents have been analysed and recorded, results should be submitted in a accurate storage format that could be compatible with the *exchange format* proposed by ICES (2010). The format proposed in this WG provides enough information to be recorded and allows the use of more than one method to estimate diet composition (e.g. Numerical, Volumetric, Gravimetric). A considerable amount of work has been made in recent years to coordinate and convert historic data into a common European stomach database in the ICES *exchange format* (see section 2.3.1; Pinnegar, 2014). A future regional sampling programme would benefit greatly by taking advantage of this work, which has already been used in several ICES working groups and ongoing projects.

References

Amend, A. S., Seifert, K. A., Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol. Ecol. 19:5555–5565.

Brander, K., Blom, G., Borges, M.F., Erzini, K., Henderson, G., Mackenzie, B., Mendes, H., Santos, A.M.P., Toresen, R. (2003). Changes in fish distribution in the eastern North Atlantic; are we seeing a coherent response to changing temperature? ICES Marine Science Symposia, 219: 261-270.

Cardoso, I., Moura, T., Mendes, H., Azevedo, M., Silva, C., Murta, A. (2015). An ecosystem approach to Portuguese mixed fisheries. ICES Journal of Marine Science, 72(9): 2618-2626.

Carreon-Martinez, L., Johnson, T. B., Ludsin, S. A., Heath, D. D. (2011). Utilization of stomach content DNA to determine diet diversity in piscivorous fishes. Journal of Fish Biology. 78, 1170–1182 doi:10.1111/j.1095-8649.2011.02925.x

Costa, P.R., Garrido, S. (2004). Domoic acid accumulation in the sardine Sardina pilchardus and its relationship to Pseudo-nitzschia diatom ingestion. Mar Ecol Prog Ser, 284: 261–268.

Dickey-Collas, M. (2014). Why the complex nature of integrated ecosystem assessments requires a flexible and adaptive approach. ICES Journal of Marine Science: Journal du Conseil: fsu027.

Dolgov, A.V., Yaragina, N.A., Orlova, E.L., Bogstad, B., Johannesen, E., and Mehl, S. (2007). 20th anniversary of the PINRO–IMR cooperation in the investigations of feeding in the Barents Sea—results and perspectives. In: Proceedings of the 12th Norwegian–Russian Symposium, Tromsø, 21–22 August 2007, pp. 44–78. IMR/PINRO Report Series, 5/2007.

EU (2008). COMMISSION REGULATION (EC) No 665/2008 of 14 July 2008 laying down detailed rules for the application of Council Regulation (EC) No 199/2008 concerning the establishment of a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy. Official Journal of the European Union, L 186/3.

EU (2008). Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008, establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Official Journal of the European Union L164, 19–40.

EU (2013). REGULATION (EU) No 1380/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC. Official Journal of the European Union, L 354/22.

EU (2016). PROPOSAL FOR A COMMISSION DIRECTIVE (EU) No…/… replacing Annex III to Directive 2008/56/EC of the European Parliament and of the Council establishing a framework for Community action in the field of marine environmental policy (Marine Strategy Framework Directive) (under review).

FAO (2008). Fisheries management. 2. The ecosystem approach to fisheries. 2.1 Best practices in ecosystem modelling for informing an ecosystem approach to fisheries. FAO Fisheries Technical Guidelines for Responsible Fisheries. No. 4, Suppl. 2, Add. 1. Rome, FAO. 2008. 78p.

ICES (1997). Report of the study group on multispecies model implementation in the Baltic. International Council for the Exploration of the Sea. ICES CM 1997/J:2.

ICES (2006). Report of the Study Group on Multispecies Assessments in the North Sea (SGMSNS). Copenhagen, 20–25 February 2006. ICES CM 2006/RMC:02. 71pp.

ICES (2007). Report of the Working Group on Multispecies Assessment Methods (WGSAM). San Sebastian, Spain, 15–19 October 2007. ICES CM 2007/RMC:08. 132pp.ICES 2010. Report of the Working Group on Multispecies Assessment Methods. ICES CM 2010/SSGSUE:05.

ICES (2012). Report of the Working Group on Multispecies Assessment Methods (WGSAM). ICES CM 2012\SSGSUE:10.ICES (2013a). Multispecies considerations for the North Sea stocks. ICES Advice 2013, Book 6, section 6.3.1, June 2013. International Council for the Exploration of the Sea, Copenhagen, Denmark.

ICES (2013b) Report of the Working Group on the Ecosystem Effects of Fishing Activities (WGECO). 1–8 May 2013 Copenhagen, Denmark. ICES CM 2013/ACOM:25

ICES (2014). Report of the Workshop to develop recommendations for potentially useful Food Web Indicators (WKFooWI) 31 March–3 April 2014 ICES Headquarters, Copenhagen, Denmark. ICES CM 2014\ACOM:48.

ICES (2015a). Report of the Working Group on Pathology and Diseases of Marine Organisms (WGPDMO), 24–28 February 2015, Helsinki, Finland. ICES CM 2015/SSGEPI:01. 126 pp.

ICES (2015b). Report of the Workshop on guidance for the review of MSFD decision descriptor 4 – food webs II (WKGMSFDD4-II), 24–25 February 2015, ICES Headquarters, Denmark. ICES CM 2015\ACOM:49. 48 pp.

ICES (2016). AORAC-SA FAO workshop: Making the ecosystem approach operational, 21-22 January, Copenhagen, DK. 55 pp.

Jennings, S., Greenstreet, S.P.R., Hill, L. Piet, G.J., Pinnegar, J.K., Warr, K.J. (2002).Long-term trends in the trophic structure of the North Sea fish community: evidence from stable-isotope analysis, size-spectra and community metrics Marine Biology (2002) 141: 1085–1097. DOI10.1007/s00227-002-0905-7.

Lefebvre, K.A., Silver, M.W., Coale, S.L., Tjeerdema, R.S. (2002). Domoic acid in planktivorous fish in relation to toxic Pseudo-nitzschia cell densities. Mar Biol 140:625–631.

M. Leray, M., Boehm, J. T., S. C. Mills, S. C., Meyer, C. P. (2012). Moorea BIOCODE barcode library as a tool for understanding predator–prey interactions: insights into the diet of common predatory coral reef fishes. Coral Reefs. 31:383–388. DOI 10.1007/s00338-011-0845-0

Levin, P. S., Fogarty, M. J., Murawski, S. A., and Fluharty, D. (2009). Integrated ecosystem assessments: developing the scientific basis for ecosystem-based management of the ocean. PLoS Biology, 7: e1000014.

Link, J. (2010). Ecosystem-based fisheries management: confronting tradeoffs, Cambridge University Press.

Link, J. S. and H. I. Browman (2014). "Integrating what? Levels of marine ecosystem-based assessment and management." ICES Journal of Marine Science: Journal du Conseil 71(5): 1170-1173.

Dunn, M.R., Szabo, A., McVeagh, M.S., Smith, P.J. (2010).The diet of deep water sharks and the benefits of using DNA identification of prey. Deep-Sea Research I 57. 923–930.

Skern-Mauritzen, M., Ottersen, G., Handegard, N. O., Huse, G., Dingsør, J.E., Stenseth, N.C., Kjesbu, O.S. (2016). Ecosystem processes are rarely included in tactical fisheries. Fish and Fisheries, 17: 165–175.

Murray, D. C., Bunce, M., Cannell, B. L., Oliver, R., Houston, J., White, N. E., et al. (2011). DNA-based faecal dietary analysis: a comparison of qPCR and High Throughput Sequencing approaches. PLoS ONE 6:e25776.

Perry, A.L., Low, P.J., Ellis, J.R., Reynolds, J.D. (2005). Climate change and distribution shifts in marine fishes. Science, 308, 1912–1915.

Pinnegar, J. K., Goñi, N., Trenkel, V. M., Arrizabalaga, H., Melle, W., Keating, J., Óskarsson, G. (2015). A new compilation of stomach content data for commercially important pelagic fish species in the northeast Atlantic. Earth Syst. Sci. Data, 7, 19–28, 2015

Pinnegar, J.K. (2014a). DAPSTOM - An Integrated Database & Portal for Fish Stomach Records. Version 4.7. Centre for Environment, Fisheries & Aquaculture Science, Lowestoft, UK. February 2014, 39pp.

Pinnegar, J.K. (2014b). Rebuilding the DAPSTOM database to enhance compatibility with ICES datasets. Final Report for EU Contract No MARE/2012/02 – SI2.632887. Centre for Environment, Fisheries & Aquaculture Science, Lowestoft, UK. November 2014, 29pp.

Plagányi, É. E. (2007). "Models for an ecosystem approach to fisheries." FAO Fisheries Technical Paper (FAO) 477.

Pompanon, F. B.; Deagle, B. E.; Symondson, W. O. C.; Brown, D. S.; Jarman, S. N.; Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. Mol. Ecol. 21:1931–1950.

Serra, I., Edwards, M., Martin, J., Genner, M. (2015). Warming shelf seas drive the subtropicalization of European pelagic fish communities. Global Change Biology (2015) 21, 144–153, doi: 10.1111/gcb.12747

SGRN 10-03. Review of needs related to surveys. 4 - 8 October 2010, Brussels, Belgium.

Stebbing ARD, Turk SMT, Wheeler A, Clarke KR (2002). Immigration of southern fish species to south-west England linked to warming of the North Atlantic (1960- 2001). Journal of the Marine Biological Association of the United Kingdom, 82, 177–180

Szoboszlai, A.I., Thayer, J.A., Wood, S.A., Sydeman, L.E., Koehn, L.E., Essington, T.E. (2014): The California current predator diet database: synthesis of common forage species. doi:10.1594/PANGAEA.834750.

Thalinger, T., Oehm, J., MAYR, H., Obwexer, N., Zeisler, C., Traugott, M. (2015). Molecular prey identification in Central European piscivores. Molecular Ecology Resources. doi: 10.1111/1755-0998.12436

Walther, Y. M., and Möllmann, C. (2014). Bringing integrated ecosystem assessments to real life: a scientific framework for ICES. ICES Journal of Marine Science: Journal du Conseil, 71: 1183-1186.

Young, J.W., Hunt, B.P., Cook, T.R., Llopiz, J.K., Hazen, E.L., Pethybridge, H.R., Ceccarelli, D., Lorrain, A., Olson, R.J., Allain, V., Menkes, C. (2015). The trophodynamics of marine top predators: Current knowledge, recent advances and challenges. Deep Sea Research Part II: Topical Studies in Oceanography, 113:170-87.

Appendix 1 - Questionnaire on existing stomach sampling

The fishPi project
Regional co-operation in fisheries data collection

WP 3: Sampling programmes for Ecosystem indicators,
small scale and recreational fisheries

<u>Questionnaire on stomach sampling programs</u>

The current information taken under the Data collection Framework (DCF) seems not to be enough to achieve the newly established goals of an ecosystem approach to fisheries management under the e.g. Marine Strategy Framework Directive (MSFD) and the reform of the Common Fisheries Policy (CFP). Understanding the full extent of the impact of a given fisheries in the ecosystem is a growing need, as a management measure for a specific fleet/species might have unexpected consequences to other fleets/species through ecological pathways. Fisheries management would improve from additional information for an appropriate assessment of fish species. After consultation with relevant end users and Member States (MS), the 'fishPi' project aims to set the guidelines for the design of a regional sampling plan for stomach contents, by identifying information gaps and developing methodologies to collect alternative data to produce fisheries ecosystem indicators and identify key stocks to monitor based on their ecological role.
The present questionnaire serves as a mean of cataloguing current coverage of stomach data for each MS, it will contribute to the process of data specification and development of regional sampling programs within this project.

Member state (MS): ______________________________________
Institutions collecting data: (list institutions that are stomach data owners)

Questions:

1. MS have or had a stomach sampling program? yes ☐ no ☐ (if yes fill table 1)
2. MS have stomach data from occasional studies? yes ☐ no ☐ (if yes fill table 2)
3. MS have stomach data in a database? yes ☐ no ☐ (if yes fill table 3)
4. Give an estimation of costs by stomach. Estimate the cost taking in consideration collection, analysis, data logging, consumables, labour _____________ €
5. Stomach data will be used for:
   i. Multispecies assessment models
   ii. MSFD indicators
   iii. Ecosystem models
   iv. Other (please specify)

Table 1. Stomach sampling program - metadata

Institution	Species covered	Sampling period	Geographic area	Sampling design 1 *	Sampling design 2**	Analysis level ***	Variables	Raw data availability

*dependent on existing surveys, dependent on biological samples from fisheries, both or dedicated design
**R - random, S - predator size selected, O - other
***H – highly detailed including full composition of diet species in numbers and weight or size, M – full composition of diet species in numbers, B – basic visual record of main prey

Table 2. Stomach sampling from occasional studies - metadata

Institution	Species covered	Sampling period	Seasonal (Y/N)	Geographic area	Variables	Raw data availability	Published (Y/N)

Table 3. Stomach database - metadata

Institution	Species covered	Data base size (N ind. sampled)	N years / years covered	Variables	Analysis level **	Data availability *

*P – public, SP – on request and under a protocol, NP – not public
**H – highly detailed including full composition of diet species in numbers and weight or size, M – full composition of diet species in numbers, B – basic visual record of main prey

6. The MS has the need for any species specific stomach sampling program? yes ☐ no ☐

If yes, please specify the species and justify (e.g. apex predator, keystone species, influential species):

ANNEX 17 WP3 A regional sampling plan for small scale fisheries and recreational fisheries

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP3 – Regional sampling programmes for fisheries and ecosystem impact data not currently collected (including stomach contents and by-catch)

Deliverable 3.3 A regional sampling plan for data collection of small scale fisheries and recreational fisheries.

Lead: AZTI, Spain

Core Team: Estanis Mugerza, Lucia Zarauz and Eider Andonegi (AZTI, Spain)
Hugo Mendes and Ana Moreno (IPMA, Portugal)
Mike Armstrong, (CEFAS, UK)
Patrik Börjesson (SLU, Sweden)

Table of Contents

1 Introduction

fishPi is a European project aiming to strengthen regional coordination in fisheries data collection, in line with the momentum towards a regional approach in fisheries management introduced by the Common Fishery Policy (CFP, EU 2013).

Small scale and recreational fisheries (SSF and RFS) are an important economic and social activity in many European inshore coastal areas. Despite their differences, these fisheries present analogies which justify a similar approach. They have reduced mobility, which makes them dependent on local and regional ecosystems, and focus their impact on coastal fish resources and habitats. Unlike large scale fisheries (LSF), official statistics are often limited for SSF, and they are missing for RFS. Data on catches and effort are therefore dependent on sampling if there are no census data, which has traditionally hampered the understanding of these fisheries, and underestimated their impacts.

SSF and RFS require methodological approaches that are different from the ones commonly used in European commercial fisheries (ICES 2010a, 2013a). This fact explains the inclusion of these fisheries in WP3.

This report is comprised of two main sections describing a regional sampling plan for the collection of SSF and RFS data, respectively. Each section follows the structure for the design and implementation of a regional scheme for data collection presented in Figure 11. They can be read separately, or be taken together. At the end of the chapter, we present three additional sections which are common for the two fisheries. The first one describes the main ideas which need to be considered to set up an effective regional coordination in the coastal area, including SSF and RFS data collection. The second one summarizes the interactions between both fisheries and legislation other than CFP. The final section summarizes the main conclusions of the report.

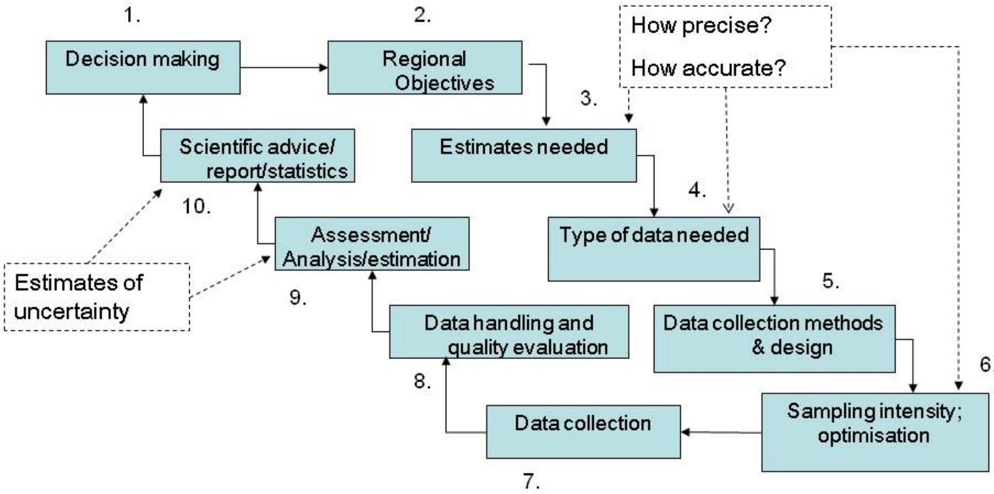


Figure 11. Regional sampling programme scheme

2 A regional sampling plan for data collection of small scale fisheries

2.1 Framework/ Introduction

This report summarizes the work undertaken under Work Package 3 (WP3) of the fishPi, to strengthen regional cooperation in small scale fisheries data collection. The report follows the structure for the design and implementation of a regional scheme for data collection presented in Figure 11. Each section of the scheme has been developed taking advantage of the work already done by the main groups dealing with small scale fisheries during the last few years, such as the Nantes workshop on "Common understanding and statistical methodologies to estimate/re-evaluate transversal data in small-scale fisheries" (Anon. 2013), ICES Working Group on Commercial Catches (WGCATCH) (ICES 2016) and, the Cyprus workshop on transversal variables (JRC 2016). Special focus has been placed on the characteristics, recommendations and challenges for regional data collection.

A key part of the work has been to contact WGCATCH as a technical advisory group on statistical and practical aspects of sampling commercial fisheries catches and hence a relevant end user, and collaborate in the preparation of the WGCATCH 2015 meeting, with a special session on SSF. In particular, a questionnaire about the dimension and relevance of SSF in Europe, and the different sampling methods used to monitor them was developed together by WGCATCH chairs and the fishPi WP3 team (Annex I). The questionnaire was sent to WGCATCH participants to be filled in, and the results were analyzed during the meeting and summarized in the WG report. This questionnaire will hereafter be refereed as the WGCATCH questionnaire.

2.2 Decision making

One of the key changes introduced in the forthcoming Multiannual Union Programme for Data Collection, Management and use of Fisheries Data (EU-MAP) is an end user driven framework, allowing for flexibility in what and how much data to collect and in methods to be applied.
There are a wide variety of potential end users in Europe that require or might require the use of SSF data, most of them having an important role within the decision making process (Table 20)

Table 20. Existing and potential end-users of SSF data.

End User	End user sub groups	Use of SSF data
ICES	ICES expert assessment WG	
	ICES WGCATCH	Documents national fishery sampling schemes, establishes best practice and guidelines on sampling and estimation procedures, and provides advice on other uses of fishery data.
	ICES WGRFS	Planning and coordination of marine recreational fishery data collection for stock assessment purposes.
	ICES PGDATA	To understand the quality of datasets and to use them as effectively as possible, and to employ objective ways to identify and prioritise data needs
Other RMFO (ICCAT, NAFO, NEAFC…)	Expert assessment and ecosystem WG	

End User	End user sub groups	Use of SSF data
European Commission	DGMARE & DG Environment	Implementation of MSFD; achievement of GES with good management of recreational as well as small scale fishery impacts. Implementation of CE 812/2004, Birds Directive, Habitats Directive
	STECF	Inclusion of data collection in the EU MAP
International Organizations	FAO, OSPAR, ASCOBANS, ACAP, IWC…	Identifying threats, recommending action plans, implementation of different agreements
Regional Coordination Groups	RCGs for each region	Coordination and cost-effectiveness of fisheries data collection within regions (if included in EU-MAP). Finding the balance between regional and national data needs
National Governments and regional fisheries authorities within countries		Developing policy positions on management that reflects the ecosystem aspects of sustainable development in coastal regions and spatial planning such as MCZs. Meeting international agreed responsibilities
Scientific community in general.	Universities; Govt. departments; other Institutes	Scientists interested on small scale fisheries and their impact in the coastal ecosystems. Interactions with other uses of the sea
		Data for publication
Representative bodies for International and national commercial fisheries.	Commercial fishermen's organisations and federations.	Policy developments in relation to small scale fisheries and their impact in the coastal ecosystems. Interactions with other uses of the sea
Recreational fisheries bodies	Recreational fishermen's organisations and federations (EAA, Angling Trust…)	Developing best practices
Advisory Councils	e.g. North Western Waters AC; North Sea AC…..	Policy developments in relation to small scale fisheries and their impact in the coastal ecosystems. Interactions with other uses of the sea
Marine NGOs	Birdlife international, WWF, GREENPEACE, OCEANA etc.	Policy developments in relation to small scale fisheries and their impact in the coastal ecosystems. Interactions with other uses of the sea

2.3 Regional Objectives and Estimates

Regional Coordination Groups (RCGs) will be the responsible bodies to define regional sampling objectives. To achieve this aim, the needs of end users (estimates, aggregation and precision levels), the feasibility of collecting such information and their potential impact on data collection programmes (available resources) will need to be taken into account.

As a starting point to define regional objectives for SSF, it is relevant to understand the drivers for the collection of SSF data, as well as the characteristics and the dimension of the target population.

2.3.1 Justification of the data collection

The main drivers for collection of small scale fisheries data can be detailed as follows:

i. SSF are of great importance in Europe, in terms of job opportunities and contribution to the economy of coastal communities (Guyader et al. 2013).

ii. SSF are highly dependent on coastal inshore waters and their reduced mobility makes them extremely dependent on local and regional ecosystem resources. Similarly, SSF have a significant impact on some coastal fish resources and essential habitats.

iii. Conflicts can arise between SSF, large scale and recreational fishing fleets; and also with other uses of the sea (e.g. renewable energies, dredging, tourism, and even conservation marine areas), because they interact in the physical surroundings.

iv. The specific characteristics, diversity and complexity of small scale fisheries, makes their assessment and management necessarily different from those applied in large-scale fisheries. Good quality information on SSF is therefore needed.

v. Despite its importance, the sector is still poorly understood, statistics are often limited and there are no clear guidelines for the design, implementation and quality assurance of sampling schemes targeting SSF. There is even a lack of a uniform and straightforward definition for SSF, which contributes to the difficulty of managing the sector and implementing targeted policies.

vi. There is a need to accurately quantify removals of fish made by SSF and to include them in stock assessments when appropriate, and to support development of long-term regional fishery management plans and the increasing focus on management of mixed fisheries and ecosystem impacts.

vii. There is a need for information on small scale fishing activity and pressures to support Marine Spatial Planning (MSP, EU 2014) and Marine Strategy Framework Directive (MSFD, EU 2008).

2.3.2 Define the target population

Diversity and specificities of SSF at the European level are extensively highlighted in the EC Study N° FISH/2005/10 and the annexes of the Nantes workshop (Anon. 2013). An analysis based on a selection of case studies (Guyader et al. 2013) showed that, as compared with large-scale fisheries, European small-scale fleets: (i) are composed of smaller vessels and, consequently , travel shorter distances to fishing grounds, and are more reliant on coastal areas, (ii) have smaller crews (although the global employment figure is similar to that of large-scale fleets in Europe); (iii) use mostly, but not exclusively passive gears; (iv) use multi-purpose fishing approaches, and can change the fish species they target during the year; (v) have smaller catches per boat; (vi) have lower total capital investments (including fishing rights), turnover and costs; and (vii) have lower fuel consumption, making them less sensitive to changing oil prices.

Despite these overall common characteristics, there is no single definition of small-scale fisheries. For the purposes of the European Maritime and Fisheries Fund (EMFF, Regulation CE 508/2014), "small scale coastal fishing" was formally defined as fishing carried out by fishing vessels of an overall length <12m and not using towed gear. In the scientific realm, several authors have proposed definitions for SSF which are based on other characteristics rather than simply on vessel size (García-Flórez et al. 2014). Ultimately, any definition is linked to the end-user needs such as stock assessment, MSP, socio-economic studies, MSFD, Marine Protected Areas (MPA), management regulation texts, etc.

fishPi agrees with the view of the Nantes workshop (Anon. 2013) on SSF which refers to fleet segments by vessel length (LOA) ranges: <10m; 10m–12m and ≥12m. However, in line with the conclusions of WGCATCH (ICES 2016), fishPi recommends that an additional 12-15m segment should be taken into account by sampling schemes targeting SSF:

- Under-10m fleet: it is considered as a separate fleet segment in relation to data collection because there is no Control Regulation obligation to supply EU logbooks for vessels under 10m.
- 10–12m fleet: it is a separate fleet segment to ensure consistency in time-series and because they are not under Vessel Monitoring System (VMS) regulation (which is critical for mapping of fishing activities for marine spatial planning or other purposes needing data at specific spatial resolution).
- 12–15m fleet: this segment is needed because many countries have put exemptions in VMS data requirement inside the 12–15 meters fleet segment so full VMS coverage of >12 m vessels cannot be assumed.

This classification of SSF allows the identification of the fleets segments which are not covered by official logbooks and/or VMS. It is important to note that due to their limited size, these fleets are also potential candidates to fall in the exemptions detailed in the Control Regulation, under which there is no obligation to register the catch in the logbook, and the landings in the sales notes, if they are below a certain quantity. These exemptions, which are further explained in section 2.5, evidence the limitations of logbooks and sales notes information to register trips with low landing volumes (usually a characteristic of SSF).

2.3.3 Description of the fisheries

A questionnaire about SSF was sent to WGCATCH members to obtain national information on numbers of vessels, fishing effort, landings weight and value for fish and shellfish for SSF. The questionnaire also asked for details on how transversal variables are collected, and how biological sampling is conducted onshore and at sea for these fisheries (Annex I, ICES 2016). The questionnaire was developed together by WGCATCH chairs and the fishPi WP3 team. Seventeen questionnaires were completed for 14 different countries. There was one answer from the Mediterranean (Malta). The rest of the questionnaires covered three regions: the North Atlantic, the North Sea and the Baltic.

The questionnaire was designed to get the overall picture by Country, and therefore the characteristics of SSF by region cannot be depicted. When reference is made to the results of the questionnaire in this report, it should be interpreted as an overall picture of the North Atlantic, the North Sea and the Baltic. Four countries were missing to complete the picture: Netherlands, Finland, Norway and Iceland. The response from Malta was not considered in this report, as the Mediterranean is already being addressed by the project “Strengthening regional cooperation in the area of fisheries data collection in the Mediterranean and Black Sea” (MARE/2014/19). The national definitions of SSF referred to vessels under 12 m in all cases but Estonia, which also included some vessels over 12m which are considered as SSF in their regulation.

Based on the SSF questionnaires and figures provided by each country, the importance of the SSF fishing sector inside European countries was addressed at WGCATCH (ICES 2016, Table 21). Here we present a summary of the main conclusions.

Under 10m and 10-12m fleet segments are of high importance in all countries in terms of number of vessels and number of active vessels (between 60% and more than 90% of the total active fleet), and consequently employment (Table 21). No small-scale fisheries are reported in Belgium and the number of active under-10m vessels in Ireland was not available. In the Basque country (ESP-AZTI) small-scale vessels under 12m represent only 20% of the total fleet but the fleet segment between 12-15 meters, which is assumed as an small scale fleet, represent 60% of the total fleet. SSF are also of important in number of trips, but they are generally less active than LSF in terms of days at sea, which can be explained by the shorter duration of their trips (generally one-day trip).

The contributions of vessels less than 12m to total landings are lower compared to other size segments. This applies for both fish and shellfish, although their contribution is higher for shellfish. The overall picture obtained for landings value is similar to the one obtained for weight, although SSF fleet segment landings have generally a higher value.

It should be noted, however, that this information may be incomplete or have different sources and quality across fleet segments (especially when independent sampling of this fleet is not carried out). Consequently, the contribution of SSF to total number of vessels, trips, landings in weight and value should assessed by fisheries, by species and by regions because significant differences could occur between them. In general, there is a risk of underestimation of SSF variables when compared to LSF with more complete sales/declarative forms (logbooks) records. It should be also noted that for countries where independent sampling of this fleet is carried out, the proportion of SSF landings compared to LSF is often higher.

Table 21. Dimension (nº of vessels), landings (tons) and value (euros*1000) of small scale fisheries for the countries/institutes answering the WGCATCH questionnaire. Year of reference is 2012. “EU vessels” refers to the vessels registered in the Community Fishing Fleet Register (http://ec.europa.eu/fisheries/fleet/index.cfm); “Licensed vessels” refers to vessels licensed to fish; “Active vessels” are those vessels displaying at least 1 landing/fishing day. Information on data sources for each value is included. Adapted from WGCATCH (ICES 2016).

Country		Vessel length classes	Number of					Landings		Value	
			EU Vessels	Licensed vessels	Active vessels	Days at sea	Trips	Fish (tons)	Shellfish (tons)	Fish (euros*1000)	Shellfish (euros*1000)
UK SCO	Values	<10 m	4951 [1]	1083	1134	46682	27911	660	11198	1109	44156
		10-12 m	420 [1]	150	150	20259	10990	23	4956	47	17615
		>=12 m	991 [1]	432	441	73104	30024	294820	53283	378548	130879
	Data source	<10 m	EU register		Buyers and sellers						
		10-12 m	EU register		EU logbooks						
		>=12 m	EU register		EU logbooks and ERS						
SWE	Values	<10 m	950	950	714	44126	43973	2524	455	7586	4214
		10-12 m	195	192	189	12313	11891	3921	467	6398	5230
		>=12 m	225	225	208	20000	12573	139830	2107	82927	23173
	Data source	<10 m	EU register		Monthly fishing journals					Sales tickets	
		10-12 m	EU register		Logbooks					Sales tickets	
		>=12 m	EU register		Logbooks					Sales tickets	
POL	Values	<10 m	470	470	460	34826	34664	8277	N/A	7865	N/A
		10-12 m	120	119	120	9367	7310	6042	N/A	5598	N/A
		>=12 m	199	199	196	21854	11358	165384	N/A	46631	N/A
	Data source	<10 m	EU register		Monthly catch reports				N/A	Calculated using avg price per kg	N/A
		10-12 m	EU register		Paper logbooks				N/A		N/A
		>=12 m	EU register		Electronic logbook				N/A		N/A
MLT	Values	<10 m	923	923	677	26922	26907	544	N/A	2995	N/A
		10-12 m	55	55	42	1589	701	218	N/A	1590	N/A
		>=12 m	82	82	65	5165	1628	1435	N/A	8087	N/A
	Data source	<10 m	EU register	ALL vessels in 2012 Fleet Register. Logbooks, CAS surveys, Market Sales Vouchers and Direct Sales vouchers are also used.		Logbooks and Catch Assessment Survey			N/A	Logbooks and Catch Assessment Survey. Market and Direct Sales vouchers are also used.	N/A
		10-12 m	EU register			Logbooks and Catch Assessment Survey			N/A		N/A
		>=12 m	EU register			Logbooks and Catch Assessment Survey			N/A		N/A
LVA	Values	<10 m	287	-	207	10891	7442	2849	N/A	1442	N/A
		10-12 m	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		>=12 m	79	-	72	8589	6271	54624	N/A	21933	N/A
	Data source	<10 m	LR register	Licensing occurs at the level of companies	Coastal Logbook	Coastal Logbook	Coastal Logbook	Coastal Logbook	N/A	Central statistical bureau Questionnaire "1-fishery" (The participation of the responders is obligatory)	N/A
		10-12 m	N/A		N/A	N/A	N/A	N/A	N/A		N/A
		>=12 m	LR register		Logbook	Logbook	Logbook	Logbook	N/A		N/A

Country		Vessel length classes	Number of					Landings		Value	
			EU Vessels	Licensed vessels	Active vessels	Days at sea	Trips	Fish (tons)	Shellfish (tons)	Fish (euros*1000)	Shellfish (euros*1000)
BEL	Values	<10 m	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		10-12 m	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		>=12 m	88	88	88	16806	4700	18638	3199	65595	10587
	Data source	<10 m	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		10-12 m	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		>=12 m	EU register		Logbooks			Salesnotes			
DNK	Values	<10 m	2168	2140	1239	39457	39118	6793	320	15737	2141
		10-12 m	134	140	130	12095	11639	6202	12999	7940	3450
		>=12 m	482	462	448	63639	35938	437608	38759	297748	68030
	Data source	<10 m	EU register	License register	Danish Fisheries Analysis Database. A combination of sales notes, logbooks and vessel register						
		10-12 m	EU register	License register	Danish Fisheries Analysis Database. A combination of sales notes, logbooks and vessel register						
		>=12 m	EU register	License register	Danish Fisheries Analysis Database. A combination of sales notes, logbooks and vessel register						
ESP IEO	Values	<10 m		4546	4017	406021	406030	3811	8037	13823	51913
		10-12 m		388	374	51134	51120	6297	1901	13659	8224
		>=12 m		1150	1114	166861	139257	208426	20849	348975	65695
	Data source	<10 m	EU register	CFPO		Sales notes					
		10-12 m	EU register	CFPO		Logbooks + sales notes				Sales notes	
		>=12 m	EU register	CFPO		Logbooks		Logbooks + sales notes		Sales notes	
EST	Values	<10 m	1608	1387	1387			8729	0	3930	0
		10-12 m	133								
		>=12 m	42	42	42	4922	3493	51923	9000	14650	29550
	Data source	<10 m	EU register	Estonian Fisheries Information System							
		10-12 m	EU register								
		>=12 m	EU register	Estonian Fisheries Information System							
UK ENG + WALES	Values	<10 m	4951 (1)	2086	2178	115652	114812	9130	20386	24413	41352
		10-12 m	420 (1)	188	192	19126	16756	3409	8679	6840	17664
		>=12 m	991 (1)	311	315	45398	20939	123273	44304	155823	72681
	Data source	<10 m	EU register		Buyers and sellers						
		10-12 m	EU register		EU logbooks						
		>=12 m	EU register		EU logbooks and ERS						
DEU	Values	<10 m	990	990	655	131630	5117	2070	N/A	4278	N/A
		10-12 m	263	263	215	20171	16404	6900	N/A	1983	N/A
		>=12 m	313	313	72	7616	4174	68250	N/A	16428	N/A
	Data source	<10 m	EU register		Landing registry	Trip summary		Landing registry	N/A	Landing registry	N/A
		10-12 m	EU register		Landing registry	Trip summary		Landing registry	N/A	Landing registry	N/A
		>=12 m	EU register		Landing registry	Trip summary		Landing registry	N/A	Landing registry	N/A

Country		Vessel length classes	Number of					Landings		Value	
			EU Vessels	Licensed vessels	Active vessels	Days at sea	Trips	Fish (tons)	Shellfish (tons)	Fish (euros*1000)	Shellfish (euros*1000)
IRL	Values	<10 m	1620	1677	Not recorded in logbooks database			3551	5578	2966	22970
		10-12 m	292	250	161	11437	10021	13466	29792	19959	53806
		>=12 m	245	280	228	31831	10152	1973247	141252	1524150	510838
	Data source	<10 m	EU register	Logbooks				Logbooks - incomplete data			
		10-12 m	EU register	Logbooks							
		>=12 m	EU register	Logbooks							
LTU	Values	<10 m	73	73	40	6142	6142	386	N/A	N/A	N/A
		10-12 m	32	32	7	352	352	138	N/A	173	N/A
		>=12 m	1	1	0	N/A	N/A	N/A	N/A	N/A	N/A
	Data source	<10 m	EU register		Monthly fishing journals				N/A	NA	N/A
		10-12 m	EU register		Logbook				N/A	Logbook	N/A
		>=12 m	EU register		Logbook				N/A	Logbook	N/A
FRA	Values	<10 m	5337	N/A	4166	368238	366929	19207	17118	N/A	N/A
		10-12 m	972	N/A	878	120280	118397	21888	25592	N/A	N/A
		>=12 m	997	N/A	933	179796	82390	262359	45818	556317	142017
	Data source	<10 m	EU register	EU register + Ifremer activity survey							
		10-12 m	EU register	EU register + Ifremer activity survey							
		>=12 m	EU register	EU register + Ifremer activity survey							
ESP AZTI	Values	<10 m	29	25	21	1609	1609	218	6	563	27
		10-12 m	21	24	20	1923	1890	289	9	859	38
		>=12 m	176	168	133	19199	8230	44620	1434	50509	6419
	Data source	<10 m	EU register	Spanish fleet reg	Azti db	Logbooks	Logbooks	Azti db			
		10-12 m	EU register	Spanish fleet reg	Azti db	Logbooks	Logbooks	Azti db			
		>=12 m	EU register	Spanish fleet reg	Azti db	Logbooks	Logbooks	Azti db			
PRT	Values	<10 m	6895	3632	3170	238775	218761	18808	1663	64295	2910
		10-12 m	316	267	252	34918	30586	11056	551	23582	1083
		>=12 m	690	596	523	100630	83369	126128	568	190361	1466
	Data source	<10 m	EU register	Fleet register	Sales notes						
		10-12 m	EU register	Fleet register	Logbook+sales notes						
		>=12 m	EU register	Fleet register	Logbook+sales notes						

(1) Note that UK SCO and GBR + WALES are combined in the EU fleet register, and therefore the same values are shown for UK SCO and GBR + WALES

2.4 Type of data needed

The Nantes Workshop (Anon. 2013) , the Planning Group on Economic Issues (PGECON, Anon. 2014) and WGCATCH (ICES 2016) noted regional differences in data collection methods and supported the need for a regional approach with more active end-user involvement to define data needs. Such a regional approach should be integrated in a general structure comprising all kinds of fisheries (LSF, SSF and RFS). The proposal is to create a list of core transversal variables as a basic requirement for vessels without logbook. The list will be extended upon request by end-users, to include additional transversal variables or more detailed levels of aggregation, as well as biological variables derived from onshore or on-board samplings schemes. Regional Coordination Groups (RCGs) should be responsible for the management of this list of core and additional variables. To this end, the RCG will review the requests for additional variables or levels of aggregation, with a focus on the feasibility of collecting such information and its possible impact on data collection programmes. The RCG will at the same time ensure the consistency of the type of data collected and data collection methods used in the same region, to ensure that information can effectively be used to fulfill end-users needs.

2.4.1 Transversal variables

The Nantes workshop reviewed the present Data Collection Framework (DCF) requirements for transversal data (Appendix VIII of EC Commission Decision 93/2010) and considered that they are too detailed for the vessels < 10 meter (<8m vessels in the Baltic), and that a broad range of this information has not been used so far. In addition, the group considered that the collection of very detailed effort variables was cost-inefficient in the case of the no logbook obligation.

The Nantes workshop defined a core set of transversal variables as a basic requirement for vessels without logbook. In 2016, the list was reviewed in the 2nd Workshop on Transversal Variables held in Cyprus (JRC 2016), which came up with a final proposal which is presented in Table 22. The Cyprus workshop highlighted that the calculation of the three proposed effort measures (number of fishing trips, Days at Sea or Fishing Days) should be in line with the methodology established for vessels carrying logbooks. However, due to the specificities of SSF, additional work is needed to devise a common methodology based on data sources other than logbooks. The collection of gear information is also tricky for SSF as they can have licenses for several gears and may use two or more gears (e.g., pots and lines or nets) in the same trip or during the same month.

Apart from the variables proposed in Table 22, the Nantes workshop and WGCATCH considered that particular attention should be given to the collection of information on the spatial distribution of fishing vessels which are currently not covered by VMS regulation. This includes all vessels under 12m, and vessels from 12-15m which are subject to VMS exemptions in several countries. Spatial information is of high interest for scientific and management purposes, as it gives information on the changes in the distribution of fishing activity resulting from control and technical measures; as well as information on the interactions between different fleets (LSF, SSF, RFS) and any other human activity competing within the same space (e.g. aquaculture farms). Techniques such as CCTV, mobile phone apps or geolocalisation tools could be useful to improve data collection; although negative attitudes within the fishing community towards these tools may present a barrier to their implementation.

Table 22. Proposed list of transversal variables to be collected for vessels without logbooks requirement (LOA<10 meters). Adapted from JRC, 2016.

Heading	Variable	Unit	Description	Coverage	Activity segmentation	Reference Period
Capacity	Number of vessels	Number	Total number of vessels	CFFR [1]	Fleet segment	
	GT	Number	Total GT of the vessels in the segment	CFFR [1]	Fleet segment	
	kW	Number	Total kW of the vessels in the segment	CFFR [1]	Fleet segment	
	Vessel Age	Number	Average AGE of the vessels in the segment	CFFR [1]	Fleet segment	
Effort	Number of trips	Number		Active vessels	Fleet segment and gear (level 3)	Quarterly
	Days at Sea	Day		Active vessels	Fleet segment and gear (level 3)	Quarterly
	Fishing Days	Day		Active vessels	Fleet segment and gear (level 3)	Quarterly
	Number of vessels	Number		Active vessels	Fleet segment and gear (level 3)	Quarterly
Landings	Value of landings totals and per species	Euro	Value of landings total and per species	Active vessels	Fleet segment and gear (level 3)	Quarterly
	Live weight of landings	Tons	Live weight of landings in kg total and per species.	Active vessels	Fleet segment and gear (level 3)	Quarterly
	Prices by species	Euro/kg	Price per kg of species landed	Active vessels	Fleet segment and gear (level 3)	Quarterly

[1] CFFR: Community Fishing Fleet Register

2.4.2 Biological variables (length, age, sex, maturity)

Biological data to be collected will depend on end-users needs, which will be prioritized at the RCG. In general, the need for biological variables depends on the type of analysis that Assessment Groups carry out for each stock (i.e. volume and length frequency of all catch fractions (landings and discards), weight, age, sex-ratio maturity, fecundity…).

However, some fleet segments have been traditionally left out of the sampling due to their low catches, on the basis that the data would have limited influence on the assessment results, and money would be better spent on sampling the fleets taking the bulk of the catches. This approach was formalized in the DCF by the approach of ranking métiers by their reported landings and sampling only the top ranked métiers taking 90% of the total. Together with the perceived difficulties of sampling SSF, this is probably the main reason for the low availability of biological data for these fleets. Our proposal is that, when prioritizing data needs and depicting a regional sampling plan for SSF, the RCG should take into account the following issues:

- Although SSF contributions to total landings are often lower compared to other size segments, underreporting of landings can give a biased view of this contribution.

- SSF operate in more coastal areas and often more sensitive habitats (e.g., nursery grounds). Catches made by SSF may present a different size structure than catches made by LSF, which can have a significant impact on the assessment.

- Although available data are scarce, it seems that SSF can contribute significantly to the overall discard rate and amount of by-catch of particular inshore species, depending on gear type and management restrictions. For some species, for example sea bass, nursery areas are located mainly or entirely in coastal waters.

- In some cases, the effect of the SSF could also be important in the by-catch of some PETS (Protected, Endangered and Threatened Species). There are documented cases for some of these PETS species, turtles and seabirds in the case of gillnets and longlines in the Mediterranean, together with harbour porpoise (*Phocoena phocoena*) in the Atlantic.

2.5 Data collection and sampling design

2.5.1 Review of data collection methodologies and data sources.

WGCATCH 2015 made an excellent review of the different data collection methodologies and data sources in relation to SSF (ICES 2016). In this section we have extracted the more relevant parts and complete them with additional information useful for the purpose of this report.

SSF present many specific features (multi-gear, multispecies fleet, high spatial distribution, high seasonality, part-time activity in some cases, direct sales, etc.) that distinguish it from the Large Scale Fleets (LSF). Hence, SSF often have to be monitored differently and specifically by a separate census or a sampling approach adapted to their special features. Data collection methods fall into two clear categories:

a) Census methods dependent on self-reporting of data by fishers, intended to have exhaustive coverage of the population (as far as is possible).

b) Sampling schemes, in which a sample from the population is randomly selected and surveyed. Two main categories of data collection methods can be distinguished:

    - Sampling schemes that use similar data reporting methods as for a census (self-reporting), but are applied to random samples of fishers.

    - Sampling schemes where the variables of interest (gear, fishing effort, fishing zones, catches etc.) are observed or surveyed directly on-site by trained survey staff (catch assessment survey) or recorded by CCTV.

Table 23 summarizes a range of possible data collection schemes for landings, effort and other transversal variables, and fleet-based biological variables such as length compositions or discards. This is not exhaustive and other schemes may exist.

Table 23. Examples of methods for data collection from small-scale commercial fisheries, for purposes of stock assessment and fishery management. Modified from WGCATCH (ICES 2016).

Type of data	Census	Sampling schemes
List of vessels by LOA	- EU register - National fleet activity database	
Spatio-temporal activity by gear type	- Exhaustive logbook - Exhaustive use of VMS or other electronic sensors. - Exhaustive sales data	- Randomized vessel intercept scheme using a site x day area frame - Randomized issue of logbooks or other recording systems using a vessel list frame - Randomized telephone survey of vessel owners - Detailed data supplied by observers - Data from CCTV, VMS or other electronic sensors fitted to samples of vessels - Data from at-sea vessel inspection by patrol vessels and/or overflight data by enforcement agencies
Catches (landings, discards and other by-catch, e.g. PETS)	- Exhaustive logbook - Exhaustive sales data	- Randomized vessel intercept scheme using a site x day area frame, with fisher interviews and direct recording of catches on board - Randomized issue of logbooks or other recording systems using a vessel list frame - Randomized telephone survey of vessel owners - CCTV cameras on random vessel selection
Length/age compositions	- Exhaustive sales data with size categories	- Port sampling scheme, e.g. using a site x day area frame - Randomized observer scheme using vessel list frame - CCTV cameras on random vessel selection

a) Census methods

In a fisheries context this usually refers to an exhaustive coverage of the population from which data are required, for example fishing vessels. An example is the EU logbook which the Control Regulation (EC) No 1224/2009 requires to be submitted for all EU registered vessels of 10m and over, recording catches and associated effort, gear and area data by day for all fishing trips. Vessels under 10m are not required to keep such logbooks, and for these vessels a sampling plan is needed unless the Member State (MS) has required such vessels to keep an EU logbook (Article 16–3) or if sales notes are supplied (Article 16–4). In the latter case, the supply of sales notes (or sales slips) and catch declarations can also be considered as census data. The use of sales notes as census data for small-scale fisheries is common practice in countries where it is mandatory for all commercial landings (irrespective of vessel size) to be sold at specific places (generally auctions within ports) and centrally registered in national databases. This register gives information on the composition in species and weight of the landings of individual vessels.

However, several exemptions and conditions in the Control Regulation result in incomplete landings data in the logbooks and sales notes for SSF, triggering a need for a sampling scheme. In particular, there is no obligation to register the catch (and discards) in the logbook if they are below 50 kg of live-weight equivalent by given species (Art. 14). And sales notes of fish products landed by vessels below 10m, or for landings not exceeding 50 kg of live weight equivalent by species, are also exempted in the cases where the MS has installed an acceptable sampling system (Art.65). Additionally, it should be noted that fishers usually group similar species under a common commercial name when filling in the logbooks and sales notes. Thus, the provided information may not allow the full discrimination of the composition in species and weight of the landings, in the case of very mixed fisheries.

Another shortfall of sales notes is that they lack information on fishing gears and effort. Despite this lack, they are broadly used as a proxy for days at sea or number of trips in the SSF, because they are often the only source of information available for vessels under 10m. The Cyprus workshop (JRC 2016) warned about the inconsistencies of such an approach, and highlighted that additional work is needed to devise a common methodology to estimate effort from data sources other than logbooks (JRC 2016).

Apart from the Control Regulation, some MS have set up systems of self-reporting forms intended to achieve an exhaustive coverage of all the population, fulfilling the coverage limitations of official logbooks and sales notes. Ad-hoc logbooks or monthly declarations, delivered to a whole fleet segment of the census (i.e. vessels <10) are an example of this kind of census. They are usually adapted to the specificities of the SSF, and can exhibit different degrees of complexity (number and type of variables requested) and periodicity (daily, weekly, monthly…). In relation to these systems, WGCATCH 2015 highlighted the need to define how national declarative forms should be designed to ensure that key variables can be collected and recorded in a consistent way for SSF within the region considered. This is important in order to merge collected data and meet end-user needs. Additionally, it has to be taken into account that self-reporting of fishing activity, landings, discards or size-frequencies of catches present particular logistic difficulties for small vessels, and the reporting burden may have to be reduced as far as possible and technological solutions found to minimize the work needed. Strong incentives may be required to ensure continued supply of data.

A general comment for the census approach is that the reliability/completeness of the data available has to be assessed for census data collection. If unreliable/incomplete data are a major issue, countries should implement a sampling approach to estimate needed variables, or to assess the quality of the data available from the census.

b) Sampling methods

Different sampling schemes could be designed for different purposes but they have to be linked with the end-users needs. For example, data needed for marine spatial planning may require a different design than data needed for stock assessment.

In LSF, sampling is usually dedicated to biological variables (length distribution, age, maturity), as their transversal variables are in principle completely covered by the Control Regulation. In contrast, for SSF there is a need to sample both transversal and biological variables. Data on transversal variables, are needed for those fleet segments which are not covered by the Control Regulation (i.e. vessels <10m), and to improve the information of those covered segments whose data quality is limited by the exemptions foreseen in the regulation. Biological variables can only be obtained through a sampling programme.

The Control Regulation requires MS to implement sampling plans for under-10m commercial vessels, but allows exemptions, based on risk analysis, which prevents the complete documentation of landings (Annexes XIX and XX of the Commission Implementing Regulation EU No 404/2011). Appendix VIII of the EU Data Collection Framework Commission Decision 2010/93/EU also places a legal requirement for MSs to collect transversal variables including vessel numbers, fishing effort, landings and landings value. The combination of the Control Regulation and Data Collection Framework requirements for providing data on effort and landings using sampling schemes, leads to a lack of clarity on what is acceptable as a sampling scheme. It is also noted that there are different purposes leading both regulations (control and scientific analysis), which may hamper the comparability of the results.

A broader insight is needed into how sampling schemes for SSF should be designed. An approach is needed which goes beyond the narrow interpretation of these two EU regulations, and is in line with the Statistically Sound Sampling Schemes (ICES 2011, ICES 2012a, ICES 2012b, ICES 2013b). Sampling schemes for SSF have similar analogies in recreational fishing where the methods have been well tested internationally (ICES 2012c). WGCATCH plans to develop an initial draft of generic and specific guidelines for good practice in collection of transversal and biological data from small-scale fisheries in Europe, and review these during its 2016 meeting. Common guidelines are essential to ensure harmonization of data collection methods and data quality needed for an effective use of the data collected regionally.

Data collection methods used in a sampling approach may be based on self-reporting made by fishers (like for census), or on observation made directly on-site by trained staff or recorded by CCTV. On-board observations on SSF are usually difficult because the number of people which a vessel can carry is limited.

A different approach is needed to collect information on the geo-localized effort made by vessels not covered by VMS regulation. These data may be collected through the implementation of technical instruments (electronic devices) collecting information on effort with high spatial resolution. The Nantes workshop (Anon. 2013) suggested the inclusion in the EU-MAP of a provision for pilot studies coordinated at regional level to propose the methodology and to assess the cost for collecting this information.

In general, two approaches are used to select the sample:

(1) A stratified sampling of vessels from a vessel list frame, which is possible thanks to the frequent availability of a more-or-less complete register of vessels with vessel details such as length. Such availability allows for the randomized selection of vessels for collection of data and also for the estimation of the number of active vessels per landings site. This is a major advantage for surveys of small-scale commercial fisheries relative to surveys of recreational fisheries. Data must be collected from all vessels randomly selected from the list, even if the vessel was not fishing. As with at-sea sampling schemes to estimate discards in LSF, recording and interpretation of non-responses and refusals is extremely important to evaluate potential for bias.

(2) A clustered sampling of fishing trips occurring on visits to landing sites (e.g. site-day sampling frame). This method consists on a sampler recording on site the landings of all or a random sample of vessels landing at a site and day, which can be raised to all sites and days in the year using the hierarchical cluster sampling probabilities at each stage (Vølstad et al. 2014, Demanèche et al. 2013). Possibilities also exist for other combinations of methods such as the use of aerial surveys to estimate effort combined with intercept surveys to record catches and other data.

Quality issues on sampling schemes are related to the statistical soundness of the sampling design (how the sample is selected), problems arising at the implementation stage (e.g., sampling departs from randomness; refusals to provide data; strata with no or inadequate samples), and errors introduced by inappropriate estimation procedures or inaccurate information used to calculate sample probabilities. If self-reporting methods are used, validation schemes are needed to validate the supplied information in terms of completeness and reliability.

2.5.2 Coverage of current sampling

Questionnaires analyzed in WGCATCH 2015 showed the different approaches used by MS to sample SSF (ICES 2016). As in section 2.3 (Dimension of the fisheries), here we present a summary of the main conclusions obtained in WGCATCH. Corresponding tables and figures can be found in the working group report (ICES 2016).

Transversal variables (Landings by species, Gears, Fishing effort data):

Questionnaires show that transversal variables are usually sampled by two different approaches: A "census" type with a declarative form and/or a "sampling" type with a statistical approach to estimate transversal data.

Most of the countries surveyed used a census approach based on adapted declarative forms such as specific logbooks, logbooks adapted to licenses or monthly declarations. Sales notes were also used by most countries.

A sampling approach to estimate landings for SSF was also carried out, with two different approaches: stratified sampling of vessels from a vessel list frame and clustered sampling of fishing trips occurring on visits to landing sites (spatial/time sampling). Catch assessment surveys at landing sites appear to be the most common approach adopted, but it should be noted that only landings can be directly observed (i.e., discards are not estimated and gear, fishing ground, effort estimates rely on additional questions asked directly of the fishermen).

In general, all Countries collect transversal variables of SSF in some way, which could be interpreted as a good sampling coverage. However, it should be noted that in many cases the completeness and reliability of collected data are not evaluated, and therefore data quality and actual coverage is difficult to assess. Census data, for example, have the risk of assuming complete coverage, and that is not always the case. Sampling data should be evaluated in terms of bias and precision and, in cases where sampling schemes are dependent on self-reporting by fishers, the supplied information should be validated.

On shore sampling for length and age data

All countries surveyed reported on-shore sampling for length and age data. As deduced from the questionnaires analyzed in WGCATCH 2015, most institutes have their SSF sampling included in a general shore sampling scheme where landings are sampled across all vessel sizes. Some institutes also have specific sampling schemes for SSF that are used for particular cases. Although SSF vessels may land their catches at many small harbours or beaches around the coast, the catch may be transported to larger harbours and auctions for sale, where they can be sampled as part of a port sampling scheme. A problem occurs for catches disposed of directly to the public at small landing sites without sales slips due to Control Regulation exemptions, if such fish persistently differ in size composition from those sent to markets.

Onboard sampling: discards and age/length sampling

The difficulty in carrying out on-board observations on SSF is often that the number of people vessels can carry is limited. In some countries it is just impossible, as they have health and safety regulations in place preventing observers from going aboard small boats, for example where there is only one fisher on board. Even for countries that include all the vessels irrespective of size in the on-board observation sampling frame, small vessels are not often selected and refusal rates are high. For these reasons, the coverage of on board sampling schemes is usually partial, including only a part of the SSF. To some extent this can be dealt with by post-stratifying the fleet and samples by vessel size but the problem of poor coverage and high refusal rates remains for the small-vessel strata. Some countries opt for setting up self-sampling programmes to overcome these difficulties.

The number of years that SSF have been sampled varies between countries and between gears within countries. However, even without comprehensive data, it is possible to see that SSF can contribute significantly to the overall quantities of discards and amount of bycatch of particular inshore species.

2.6 Data collection

As shown in section 2.5.2 (Coverage of the current sampling), all countries collect information on SSF in some way, through a census or a sampling approach. However, in many cases the completeness and reliability of collected data are not evaluated, and therefore data quality and actual coverage is difficult to assess. There is a need to harmonise the methodologies for the design, implementation and data quality evaluation of data collection programmes, to ensure that collected data can be merged and used at a regional level to fulfill end-user needs.

WGCATCH plan to do intersessional work to develop an initial draft of generic and specific guidelines for good practice in collection of transversal and biological data from small-scale fisheries in Europe, and review these during its 2016 meeting.

The description of existing data collection methodologies for SSF, is a first step for any harmonization exercise. These should include:

- The data collection methodology carried out, which may be based on a census or in a sampling scheme. In the case of sampling schemes, the guidelines developed by WKPICS (ICES 2013b) and WGCATCH 2014 (ICES 2015a) may be a good starting point to ensure that the schemes are based on sound statistical design. The description of the sampling scheme should include: the sampling frame and its coverage of the population; the sampling units and how they are stratified in the frame; how the sampling units are selected; data that are collected from each unit; ancillary data that are collected, including refusal and non-response rates; how they are archived and quality assured; analysis methods used including quality indicators.
- How national declarative forms (ad-hoc logbooks) are designed, to ensure that key variables can be collected and recorded in a consistent way for SSF within the region considered.
- The temporal coverage of the data collection (annual, biennial, etc.).
- The quality assurance framework.

2.7 Data management, analysis and quality evaluation

2.7.1 Data management

Data management includes the archiving of data and the processes of quality assurance and quality control.
A distinction should be made between transversal variables collected under the Control Regulation (official logbooks and sales notes); and the equivalent data collected using an ad-hoc census or a sampling approach.

Official logbooks and sales notes supplied in accordance with the EU Control Regulation are stored in national data bases, and also in Regional Data Bases (RDB) when they are in place (i.e. FishFrame).

For vessels supplying such data as specified by the Control Regulation, the standardization of the data allows the merging of anonymised data coming from different MS in a single data base, although even in this case there can be national differences in the way data are coded. Data collected using an independent census or a sampling approach, however, is only stored in national data bases. The lack of harmonization in the way such data are recorded and coded in national data bases will hamper the merging of national data, and existing RDB are not designed for their storage. Addressing this will be a key task for development of a regionally coordinated sampling programme for SSF in which data collected nationally under the Control Regulation and Data Collection Framework can be accessed by Regional Coordination Groups and other end users according to permissible levels of data aggregation.

Thus, we are in a situation where the information on SSF stored in the current RDB is not exhaustive, due to the limitations of official logbooks and sales notes. There is no clear way to identify data gaps and coverage deficiencies, and those countries with data collection schemes for this segment independent of Control Regulation data, which are able to give more reliable estimates, cannot include that information into the current RDB. Thus, it becomes clear that there is a need for a regional data base adapted to the specificities of SSF. The following data sources need to be taken into account

There is a need for a common data base adapted to the specifities of SSF, which allows incorporating all available information. The following data sources need to be taken into account:

a) Official logbooks and sales notes for SSF, which are already stored in the RDB.

b) Census methods dependent on self-reporting of data by fishers, intended to have exhaustive coverage of the population. The structure of the RDB for SSF may be similar to the current one using EU logbooks and sales notes, but the required variables need to be revised.

c) Sampling schemes to estimate transversal variables such as effort and catches. Such schemes may involve probability-based sampling from a list frame of vessels or from an area frame of site-day shore sampling units. Data are obtained either by direct observations, by interviews with skippers on completion of their fishing trip or by providing logbooks to the selected skippers to record data over a longer period.

   As mentioned before, sampling schemes for SSF have similar analogies in recreational fishing (i.e. the need to estimate transversal variables) and there would be commonalities to the design of a RDB for archiving and analysis of data from both types of fisheries. For example, where recreational fishers have to be licensed, the license list is analogous to the vessel register for commercial vessels though both may have coverage issues to be addressed. It may be the case that a SSF present no or very incomplete vessel register. In this case, estimation for the total population of vessels is needed. It may be necessary to quantify the fleet and effort using site visits and possibly aerial surveys or at-sea transect surveys on vessels if feasible, an exercise which increases in difficulty as the length and complexity of the coastline (e.g. Norway) and the size and diversity of the fleet increases. The design of a data base for hosting these data may be different that of the existing RDB. Further work is needed to determine the needs and specifities of such a data base.

d) Sampling schemes to estimate biological variables (length, age etc.). As with large scale fleets, the requirement is to collect representative samples and have knowledge of the sampling probability so that samples can be raised correctly to the total catches within strata, using the hierarchical cluster sampling probabilities at each stage.

It is clear that, in the absence of exhaustive logbook schemes, a wide range of sampling schemes are possible for SSF depending on the characteristics of the fishery and the size and nature of the coastline. Data could come from schemes as diverse as aerial surveys and fisher self-sampling. However, the underlying structure of SSF and their catches is no different from large scale fisheries, involving vessels, trips, hauls, catches by species by haul, and disposal of the catches either as discards or as landings. Provided data are collected by probability-based sampling, and the probabilities are known, the analysis methods are relatively straightforward using the types of routines available in the R-survey package.

2.7.2 Quality evaluation

Data collected for SSF are intended to serve different purposes (stock assessment, marine spatial planning, marine strategy framework etc.). End users are interested in a relatively high-level overview of data quality, particularly the precision, the potential level and impact of bias, and how quality varies between the countries providing the data.

With multiple countries contributing to these estimates, one of the roles of the future Regional Coordination Groups will be to ensure these combined estimates can be quality assured. To this end, a Quality Assurance Framework (QAF) for documenting and archiving data quality is required at a national and regional level. Quality evaluation procedures for small scale fisheries data would follow the same principles as for LSF and recreational fisheries:

- Evaluation of survey design and analysis methods against guidelines for good practice documented by ICES expert groups and EuroStat.
- Documentation of quality issues at implementation stage, e.g. refusal and non-response; coverage problems, and evaluation of potential for bias arising from these.
- Quality control of archived data.
- Development of quality indicators related to bias and precision.
- Peer review of sampling survey designs.

For many of these tasks, quality evaluation of small scale fisheries data can benefit from the outcomes of WP4.

In general, several quality issues can be highlighted depending on how data are collected (through a census or a sampling scheme).

Census methods
Census methods are dependent on self-reporting of data by fishers, and are intended to have exhaustive coverage of the population (as far as is possible). Quality issues are related to actual coverage of the scheme, response rates, reliability and accuracy of data.

Like in all methods based on self-reporting, quality assurance schemes are needed to validate the supplied information, in terms of reliability and completeness. The use of technology such as CCTV can be used to validate that reported data matches what was caught, and may even allow independent estimates to be derived. If unreliable/incomplete data are a major issue, countries should implement a sampling approach to estimate needed variables, or to assess the quality of the data available from the census.

Regional guidelines need to assess countries and institutes about the methodology to validate self-reporting information. They also need to define how ad-hoc declarative forms are designed to ensure that key variables can be collected and recorded in a consistent way for SSF within the region considered.

Sampling schemes

Sampling schemes are based in the random selection of a sample from the population, which is surveyed and extrapolated to the total. In general, quality issues in sampling schemes are related to the statistical soundness of the sampling design (how the sample is selected), problems arising at the implementation stage (e.g., sampling departs from randomness; refusals to provide data; strata with no or inadequate samples), and errors introduced by inappropriate estimation procedures or inaccurate information used to calculate sample probabilities. For those schemes based on self-reporting, accuracy of the data is also an issue, and the reliability/completeness of the data collected has to be assessed (as for the census methods).

Regional guidelines will provide the basis for the design and implementation of appropriate sampling schemes to survey SSF, taking into account their special features. They will be in line with the work made by WKPICS2 (ICES 2013b), to develop general guidelines for best practice in catch sampling schemes, and quality indicators to identify bias and provide indexes of achieved precision.

2.7.3 Peer review of sampling survey designs

Designs of surveys of SSF and recreational fisheries can be relatively complex with considerable potential for bias due to inappropriate methods. The failure of a survey to provide data of the required quality can be expensive in financial terms, as well as impacting the quality of stock assessments and advice. Expertise in survey design has been building within the ICES community in Europe since the late 2000s. However, there is a need for new survey programmes to be peer-reviewed by professional statisticians with expertise in such surveys. In the USA, marine recreational fisheries surveys have been running since the 1980s and were designed with extensive input from professional statisticians. The Marine Recreational Fisheries Statistics Program (MRFSS) underwent a major peer review by the US National Academies of Sciences, Engineering, and Medicine in 2006, leading to a revised Marine Recreational Information Program (MRIP)[3]. This level of design and scrutiny of fisheries sampling schemes in Europe has seldom been achieved but is increasingly needed as surveys become developed that are based on rigorous statistical principles so that the data can be used with confidence.

[3] http://www.st.nmfs.noaa.gov/recreational-fisheries/index

3 A regional sampling plan for data collection of recreational fisheries

3.1 Framework/ Introduction

This report summarizes the work done in the frame of fishPi project WP3, to strengthen regional cooperation in recreational fisheries data collection. The report follows the structure for the design and implementation of a regional scheme for data collection presented in Figure 11. Each section of the scheme has been developed taking advantage of the work already done by the main groups dealing with recreational fisheries during the last few years, principally the ICES Working Group on Recreational Fisheries (ICES 2012c, ICES 2015b). Special focus has been placed on the characteristics, recommendations and challenges for a regional data collection.

3.2 Decision making

For fish stocks where recreational fishing is a significant source of fishing mortality, a lack of estimates of recreational catches for inclusion in stock assessments may lead to significant bias in the assessment results and provision of incorrect advice on fisheries management. Also, where there is a need to control exploitation, information is needed on the characteristics of the fisheries, their selectivity patterns and fishing behavior so that control measures can be well designed and can be enforced. Data collection on recreational fisheries is therefore needed for this purpose, as well as to monitor the effectiveness of the measures, for example minimum landing sizes or bag limits. Without this information, the ability to manage effectively may be severely impeded. Data on sea bass landings from recreational surveys in the UK, France, and the Netherlands are incorporated in ICES stock assessments (e.g. ICES 2014a, ICES 2014b) and were used by the European Commission in 2014 to develop proposals for increased size limits and bag limits for recreational catches alongside measures applied to commercial fisheries. Recreational harvests from this stock in 2012 were estimated to be around 25% of the total fishery landings. Other than Atlantic salmon, there are few other European examples of recreational fishery survey data being included in stock assessments and fishery management. Examples include Norwegian coastal cod (around 33% of total landings) and western Baltic cod (around 25% of total landings), see ICES advice at "http://www.ices.dk/community/advisory-process/Pages/Latest-advice.aspx".

There may be many other species where recreational fishing has a significant impact but no information is available. This may include inshore stocks targeted mainly by recreational and small-scale commercial fisheries.

Decision making on recreational fisheries therefore includes decisions at an international level for management of stocks under the Common Fisheries Policy, where recreational fisheries have a significant impact, as well as decisions at a national level for management of inshore stocks harvested exclusively or mainly by one country. Knowledge of recreational fisheries activities and their impacts for a wide range of species is also needed for development of spatial measures such as marine protected areas.

There are a wide variety of potential end users in Europe that require or might require the use of SSF data, most of them having an important role within the decision making process (Table 24**Error! Reference source not found.**).

Table 24. Existing and potential end-users of recreational fishery data (ICES 2015b).

End-user	End-user subgroups	Use of data
ICES	Working Group on Recreational Fishery Surveys (WGRFS)	Collation of participation, catch and economic data by country and area; Quality assurance of data collected; Development of survey methods; Provision of advice on data collection and use of recreational fishing data in stock assessment.
	Working Group on North Atlantic Salmon (WGNAS); Assessment WG on Baltic Salmon and trout (WGBAST)	Recreational catch data used in assessments
	Baltic Fish Assessment WG (WGBFAS)	Recreational catch estimates included in Western Baltic cod assessment; recreational flounder catches considered by WGRFS as suitable for assessment.
	Working Group on Celtic Seas Ecoregion (WGCSE)	Recreational catch estimates for sea bass used in assessment.
	Working Group on eels (WGEEL)	Recreational catch data sought but not sufficient for use in assessments
	Other assessment Working Groups, and Expert Groups / Steering Groups dealing with ecosystems assessments	Recreational catches of all species other than salmonids, bass, Baltic cod are needed to more completely evaluate human impacts on ecosystems and for single species assessments for stocks where recreational harvests are significant contributor to fishing mortality.
NASCO	Working groups dealing with salmon	Recreational catch data used in assessments
European Commission	DG MARE	Recreational survey data used by Commission in 2014/15 to review effects of MLS and bag limits for sea bass management. Future requests may be envisaged for other species.
	DG Environment	Implementation of MSFD; achievement of GES with good management of recreational as well as commercial fishery impacts.
Regional Coordination Groups	RCGs for each region	Coordination and cost-effectiveness of national recreational fishery data collection within regions
PGECON		Evaluating social and economic impacts of fishing and relative value of commercial and recreational sectors. This may be best done by occasional one-off surveys than in annual surveys, so we may not want to consider this as a recurrent EUMAP requirement.
National Governments and regional fisheries authorities within countries		Developing policy positions on management that includes controls on recreational fishing and aspects of sustainable development in coastal regions. Management of recreational fishing in context of spatial planning such as MCZs.
International and National recreational fishing bodies	European Anglers Alliance; national marine recreational fishing bodies, etc.	Developing policy and lobbying positions on management and sustainable development of marine recreational fishing.
National and local businesses	Charter boat businesses; tackle trade; boat manufacturers; hotels etc.	Time-series of effort and catches by species and region are useful for planning, and local authorities could benefit when making decisions on local development if they have data on how much recreational fishing takes place and the economic

End-user	End-user subgroups	Use of data
		value.
Scientific community in general.	Universities; Govt. departments; other Institutes	Scientists working on impacts of climate change should be interested in how recreational fishery species compositions are changing in each region and occurrences of species beyond previous range. Development of new recreational survey methods requires evaluation of data from existing surveys. Data for publication
Journalists	All media	Information for media articles on news items referring to recreational fishing.
Representative bodies for International and national commercial fisheries.	Commercial fishermen's organisations and federations.	Policy developments;
Advisory Councils	e.g. North Western Waters AC; North Sea AC, etc.	Policy developments
Marine NGOs		Policy developments

3.3 Regional objectives and Estimates

Recreational fisheries are of great importance in Europe, both as a pastime but also in terms of contribution to the economy particularly in coastal communities (Pawson et al. 2007; Armstrong et al. 2013). Sea angling is a major component, but in some areas recreational fishing using spears or commercial-type gears such as gillnets, long-lines and pots also takes place. Recreational fisheries are highly dependent on coastal inshore waters and their reduced mobility makes them extremely dependent on local and regional ecosystem resources where they may have a significant impact on some coastal fish resources. Conflicts can arise between SSF, large scale and recreational fishing fleets; and also with other uses of the sea (e.g., renewable energy developments, dredging, tourism, and even conservation marine areas), because they interact in the physical surroundings.

There are many drivers for the collection of recreational fishery data, and there is a wide range of existing or potential end users of the data (Table 24). The view of WGRFS (ICES 2015b) is that the following are the main drivers for collection of recreational fisheries data:

i) The need to quantify the catches and fishing mortality exerted on stocks by recreational fishing, and hence identify stocks where this needs to be included in assessments and advice.

ii) The requirement for information on recreational fisheries to help design effective and enforceable control measures where these are needed, and to monitor the outcomes.

iii) The need for a comprehensive description of regional and often poorly monitored small-scale fisheries, which include large numbers of small commercial and recreational vessels using similar methods and targeting similar species and areas, and for an evaluation of their impacts on stocks and their social and economic benefits to coastal communities to help develop local management of shared resources.

iv) The need for information to support development of long-term regional fishery management plans and the increasing focus on management of mixed fisheries and ecosystem impacts.

v) The need for information on recreational fishing activity and pressures to support marine spatial planning.

The specific objectives for a recreational fishery survey or set of surveys for a fish stock or assemblage of stocks in a region will depend on how many of these drivers are being addressed. Examples could include:

- For the purposes of stock assessment, the most basic objective is to obtain an estimate of total harvest of the stock being assessed, with associated quality indicators related to precision and bias. The current EU Data Collection Framework includes a requirement to estimate harvests of Atlantic salmon, freshwater eels, cod, sea bass, bluefin tuna and sharks (effectively all elasmobranch species) with regional differences in species that are included.

- To help in design of management measures such as minimum conservation reference sizes and bag limits, and monitoring their effectiveness, more detailed information would be required from surveys including size compositions of retained and released fish, and catch rates per day. The survey objectives would have to include collection of such data.

- To support marine spatial planning, survey objectives would include mapping of recreational fishing activities at the required resolution.

3.4 Type of data needed

The basic type of data needed from recreational fishery surveys depends on the survey methods adopted. In general data are collected on the population of recreational fishers (number of people; fishing effort; areas fished; fishing methods used etc.); catches (retained and released); and size compositions. Economic variables are also important for this fleet due to its impacts in coastal communities. Additional information such as the demographic characteristics of the population of recreational fishers and any people recruited into panels to collect catch data, are required for particular methods. This is explained in the following section.

3.5 Data collection and sampling design

3.5.1 Review of data collection methods and data sources.

ICES WGRFS reports contain detailed information on data collection methods for recreational fishery surveys, and descriptions of surveys currently being carried out in Europe (ICES 2010a, ICES 2012c, ICES 2013a, ICES 2015b). In general these cluster into off-site surveys (e.g. postal surveys of a population; use of panels of fishers to keep catch diaries), and on-site surveys where fishers are intercepted on their fishing trips. The types of data needed, and the methods available to collect these data, are summarized below.

Numbers of recreational fishers and their fishing effort

The ability to quantify the population of recreational fishers and their effort for different fishing methods is easiest when there are lists of recreational fishers (e.g. from licensing) or of recreational fishing boats (e.g. lists of registered charter boats or for-hire boats). If such lists exist, they can be randomly sampled with known probability to collect information on fishing effort, catches or other information. This approach is used in the UK for estimating Atlantic salmon catches from license lists. It is also the case of Spain where it is mandatory by regulation to be an owner of a fishing license to practice any recreational fishery modality.

In many cases, such lists are not available. In such cases, other approaches are possible:

1. Carry out probability-based nationwide surveys of the population of a country to estimate the numbers of people who fished recreationally during the year and to collect other demographic and avidity (frequency of fishing) information needed for subsequent analysis. Methods that have been used include postal surveys and random digit dialing surveys. Respondents can provide recall data on previous effort and catches, but this is subject to recall bias. A less biased approach is to recruit panels of fishers to keep diaries recording fishing trips and catches, and use the estimated total number of fishers in the country from the nationwide survey to raise the catch data from the diary panel to the whole population of fishers.

2. Direct observations at randomly selected stretches of coastline on randomly selected days where all fishing activity is recorded. Data from the observed sites is raised to all sites and days. This is extremely difficult if done on land, unless the coastline is very small and easily accessible. Other approaches with better coverage include aerial surveys. Transect surveys on vessels can also be used to count recreational fishing boats at sea at randomly selected areas and times.

Examples of these approaches can be found in the reports of the ICES Working Group on Recreational Fisheries.

Catches, retained and released fish
Total recreational fishery harvests include fish that are retained, and those that are returned to the sea but do not survive. Estimation of total harvests therefore require data on numbers (and weight, if harvests are to be given by weight) of fish retained and released, and data from experiments estimating the survival rate of released fish. The ICES WGRFS has highlighted the need for such data, and has reviewed studies on post release mortality and factors affecting mortality (ICES 2015b). Catch data can be recorded by several methods:

1. Where list frames of fishers or recreational fishing boats are available, these can be randomly sampled in order to interview the fishers or skippers to recall their catches or to request the completion of catch records for a defined period.

2. Collection of catch data from panels of fishers recruited from nationwide surveys. The mean catch per fisher in the panel is raised to the total number of fishers from the nationwide survey. The catches by panel members can be recalled for a previous period (subject to recall bias), or from records kept over a period of time using catch diaries.

3. On-site intercepts of fishers either while they are fishing (roving creel surveys) or whilst they are leaving the site on completion of fishing. This method collects catch data by trip, and requires a separate off-site nation-wide population estimate of numbers of fishing trips with the appropriate stratification (e.g. from postal or phone surveys) to raise the on-site estimates to total national catches. The on-site survey is based on random-stratified selection from an area list of fishing sites. On-site surveys also allow the most accurate collection of length data for retained fish, or if released fish are observed during roving creel surveys.

Biological variables (length, age, sex, maturity)
Biological data to be collected will depend on end-users needs, which can be prioritized at the RCG or by individual countries according to their needs. In general, the need for biological variables depends on the type of analysis that Assessment Groups carry out for each stock. If catch weights are required, information on length compositions will be needed to convert numbers caught to weight caught. Length data are also needed to determine the selectivity of the fishery and the release rate

as a function of length, or if length or age compositions of the fishery are required as input to an assessment model. If post-release mortality varies with fish size, length data will be needed to estimate total release mortality. Collection of representative length data can be extremely difficult in recreational surveys, particularly where data are recalled or recorded in diaries. In the Netherlands, significant differences were found in sea bass length compositions from diary records and on-site observations (Van der Hammen et al. 2012). To facilitate combination of national estimates, all length compositions should be collected in the same length bins.
Currently, the DCF does not specify a requirement to collect size or age data for recreational catches of stocks specified in the Commission Decision.

Economic variables
The DCF does not require collection of economic data for recreational fisheries. However, this has been done in previous recreational fishery surveys by asking fishers to record spending on different items for a fishing trip or over a period of time. Based on such an approach, Armstrong et al (2013) estimated that sea anglers in England spent £1.23 billion on their sport in 2012, which, taking indirect and induced effects into account, supported £2.1 billion of total spending and a total of over 23,600 jobs. There are however a number of different ways to evaluate the economic value of recreational fishing, and in the event of any regional requirement to make such estimates, a strict harmonization of methods would be required.

Regional harmonisation of methods
There is a range of methods for estimation of recreational catches, and the choice of method will depend on the specific conditions in a country as well as the resources and expertise available. For a regionally coordinated recreational fishery survey programme involving surveys in several countries, it is not necessary to have the same survey design in each country. The most critical requirement is that each survey is designed along statistical sound principles to minimise bias and to allow reliable estimation of precision. For example, it could be possible to have sampling from a license list frame in one country, a nationwide survey and diary survey in another country, and a mixture of off-site population survey and on-site intercept survey in a third country. Each survey provides catch estimates with quality indicators of precision and bias, and the catches from each survey can be summed for a given stock and the combined precision calculated. This has implications for any consideration of a regional data base for recreational fishery survey data.

3.6 Data collection

The logistics and timing of data collection varies between survey methods for recreational fisheries. Nationwide postal or telephone surveys are typically carried out in several waves during the year, whilst diary panels and on-site surveys involve more or less continuous data recording. For a regionally coordinated survey programme, this means that key periods for review of survey progress, achievements and issues arising may differ between countries.

Recreational fishery surveys have relatively high potential for bias during the implementation stage, even if the survey is designed using strict statistical protocols. Self-reporting of catches by panels of diarists may be inaccurate, and this is likely to be exacerbated if there is poor level of contact between the survey operators and the panelists. For such surveys it is important that the protocols for these contacts are very well defined, documented and monitored in each country carrying out the surveys. Methods for handling drop-outs and replacements in panels also need to be well established. Non-response and refusal rates are another potential source of bias, and these must be fully documented, and the potential for bias investigated.

For on-site surveys involving stratified random selection of sites and days, or times of day, departures from the sampling protocols can be a source of bias, requiring the performance of the sampling team to be monitored to ensure that sampling is representative. For a regional coordinated survey programme, the coordination team (under the umbrella of the RCGs and also intersessionally) must ensure that all contributing countries work to the same protocols for minimizing bias and for documenting issues that arise during implementation that could lead to bias.

3.7 Data management, analysis and quality evaluation

3.7.1 Data management

Data management includes the archiving of data and the processes of quality assurance and quality control. In a regionally coordinated programme of recreational fishery surveys, the QA/QC protocols would have to be defined and implemented in the same way as for commercial fishery sampling data, and as demonstrated by the fishPi project. An additional range of quality checks would however be needed as some of the recreational surveys are very different from any commercial fishery surveys, for example nationwide population surveys. Where these involve addition of questions on recreational fishing to an existing nationwide survey designed for other purposes (for example the UK Office of National Statistics Survey used by Armstrong et al. in 2013 to estimate sea angling participation and effort in 2012), the data management protocols will have been established by the survey provider.

The structure of recreational fishery data is analogous to commercial fishing in involving fishers, fishing trips, catches (retained & released); size compositions in a hierarchical cluster pattern, for which the sampling probabilities at each stage are needed. In the case of a list frame of recreational charter boats, for example, where vessels are selected at random and record data on catches and catch compositions, the data are directly analogous to selecting commercial vessels for at-sea sampling to estimate discards. Such data could be archived and analysed using the same approaches. Selection of anglers from a license list frame results in a similar data structure. A difference arises where the selected vessels or anglers are required to keep a catch diary or log book for an extended period. For example, Armstrong et al (2013) established a vessel/month frame for charter boats and selected vessel-month combinations for completion of catch logs (on the basis that longer periods cause data collection fatigue). In that case the primary sampling unit (PSU) is a vessel/month rather than a vessel trip as in discard observer schemes. If annual diaries are kept by individual fishers on a panel, the PSU is the fisher.

For on-site surveys, particularly roving-creel surveys, the data structures are more complex. The PSUs are typically sites/days or sites/days/periods of day. In a roving creel survey, incomplete trips are observed, and additional data are needed to expand observed data to complete trips, and to account for biases such as length-of-stay bias (fishers on site for short periods are less likely to be observed).

For commercial large-scale fisheries, the data needed for raising sample estimates to the fleet as a whole are typically transversal variables from exhaustive EU logsheets and sales notes. The detailed transversal variables are held in a separate data base from the Regional Data Base for sampling data, and are imported at appropriate level of aggregation for raising the data in the RDB. For small scale commercial fisheries, the transversal data are often incomplete and need to be estimated from surveys. This is closer to the situation for recreational fisheries, but the additional surveys for recreational fisheries may be of a very different nature, such as nationwide postal or telephone surveys, the data for which will be held in bespoke data bases set up for those surveys. These surveys also provide data such as age profiles, fishing avidity and home location that are needed for post-stratification of the population survey and catch sampling surveys to re-weight the catch sample data to be representative of the population. The concept of an RDB for recreational fishery data is therefore more complex than for commercial fisheries and a principal focus at this stage should be to ensure that data from national surveys of different types are properly archived and subject to appropriate QA/QC procedures, and that the full process from survey design, implementation, data archiving and quality control, data analysis and reporting is fully documented and transparent for each country contributing to a regionally coordinated recreational survey programme.

3.7.2 Quality evaluation

Quality evaluation procedures for recreational fisheries data would follow the same principles as for large scale and small scale commercial fisheries:

- Evaluation of survey design and analysis methods against guidelines for good practice documented by ICES expert groups and EuroStat.
- Documentation of quality issues at implementation stage e.g. refusal and non-response; coverage problems, and evaluation of potential for bias arising from these.
- Quality control of archived data.
- Development of quality indicators related to bias and precision.

As mentioned above, a large variety of methods is available for surveying recreational fisheries. Different approaches have their own strengths and weaknesses, and are more or less appropriate according to the scale and objectives of each particular survey (Pollock et al., 1994; ICES 2010a). In general, we can distinguish between off-site methods, in which fishers are surveyed after fishing activity has occurred (i.e. phone, mail, and diaries), and on-site methods, in which fishers are interviewed during or immediately after fishing, at locations near the fishing activity (i.e. aerial, access point, and roving surveys).

Off-site surveys can be more cost-effective and accessible, and they are used to collect information on recreational effort and harvest in many European MSs (ICES 2010a). Their main drawback is that they are known to be associated with several biases, of which coverage, non-response, and recall biases are the most dominant (Tarrant et al., 1993; Connelly and Brown, 1995; Lyle et al., 2002; Vaske et al., 2003).

On-site methods are characterized for better quality data, as the information is directly observed in the field. However, they are usually more expensive and it may be difficult to achieve a sufficient coverage and to sample a representative portion of the total population.

3.7.3 Peer review of sampling survey designs.

As mentioned for SSF, designs of recreational fisheries surveys can be relatively complex with considerable potential for bias due to inappropriate methods. The failure of a survey to provide data of the required quality can be expensive in financial terms, as well as impacting the quality of stock assessments and advice. Expertise in survey design has been building within the ICES community in Europe since the late 2000s. However, there is a need for new survey programmes to be peer-reviewed by professional statisticians with expertise in such surveys. In the USA, marine recreational fisheries surveys have been running since the 1980s and were designed with extensive input from professional statisticians. The Marine Recreational Fisheries Statistics Program (MRFSS) underwent a major peer review by the US National Academies of Sciences, Engineering, and Medicine in 2006, leading to a revised Marine Recreational Information Program (MRIP)[4]. This level of design and

[4] http://www.st.nmfs.noaa.gov/recreational-fisheries/index

scrutiny of fisheries sampling schemes in Europe has seldom been achieved but is increasingly needed as surveys become developed that are based on rigorous statistical principles so that the data can be used with confidence.

3.8 Case study: sea bass in ICES Area IV and VII

Sea bass in the North Sea, Channel, Celtic Sea and Irish Sea is one of few examples in Europe where recreational fishery catch estimates are included in an ICES stock assessment for a marine species. A description of the surveys and existing catch estimates can be found in ICES stock assessment reports (ICES 2014a, ICES 2014b). The catch estimates (Table 25) are primarily from three national surveys (an additional estimate from Belgium was supplied informally to ICES):

1. Surveys in France involving nationwide randomized screening surveys by random digit dialing followed by selection of a panel of people to complete catch diaries.

2. Surveys in England in 2012 involving a nationwide randomized face-to-face survey (Office of National Statistics opinions survey, with angling questions added) to estimate sea angling effort, and on-site roving creel surveys to estimate catch per unit effort of shore and private boat anglers. A separate charter boat survey was carried out involving random selection of vessels on random months from a list frame of vessels, to keep a catch log book for 1 month.

3. Biennial surveys in the Netherlands involving nationwide screening survey followed by recruitment of a diary panel using a marketing survey.

The information in Table 25 shows that the estimates available to ICES cover different years, and in the case of France, the surveys in 2011-12 provided estimates for the North Atlantic as a whole (i.e. including Area VIII) and could not be disaggregated to give estimates for areas IV and VII. The methods used by each country differed, with on-site surveys used extensively by England and diary surveys by France and Netherlands.

This case example illustrates that the timing and coverage of surveys differed between countries. Each country designed its surveys completely independently, to meet DCF requirements, and there was no attempt to provide a coherent set of estimates for an individual year. The survey coverage was also incomplete for example the UK surveys excluded Wales where important recreational fisheries for sea bass occur. An inability to disaggregate the 2011/12 French data by stock area was also an impediment for the stock assessment. Had these surveys been fully regionally coordinated, the following process could have been followed, leading to a coherent set of fully documented and peer reviewed estimates of sea bass catches for use in assessments:

- Document the objectives of the surveys i.e. the estimates that are required.

- Develop a description of the recreational fisheries in each country based on existing surveys and expert knowledge.

- Document any existing estimates of recreational catches to demonstrate the potential contribution of recreational fishing to total F. Identify as far as possible the coverage, quantity and quality of recreational catch data (retained and released) that is needed to meet the objectives.

- Identify the data to be collected from the survey including any additional data required from surveys to meet national or regional requirements, for example estimation of multispecies catches.

- Identify the most appropriate and cost-effective methods for estimating recreational catches in each country, with focus on reducing bias so that national estimates can be combined reliably, and the correct variance of the combined estimate calculated.

- Evaluate the survey intensity needed in each country to deliver regional catch estimates with sufficient precision to support the assessment and advice for the stock. This could be based on results of pilot studies or surveys in other countries indicating the precision obtainable from nationwide screening surveys and the variability of catch estimates between anglers or (for on-site surveys) fishing sites and days. The size of the national survey will also be dependent on the need for estimates of recreational catches for a wider range of species, and the funding available.

- Coordinate the timing of surveys between countries. For example ICES requires catch estimates for calendar years, and the surveys should be timed accordingly.

- Develop a system to monitor the design, implementation, analysis and reporting of the international surveys – this could be a role for the lead scientists for each national survey working together through WGRFS and liaising with the RCGs. Individual and combined survey results (catch estimates and quality indicators) should ideally be reviewed by WGRFS before submission to the stock assessment expert groups.

- Submit the proposed coordinated regional programme and methods to independent experts on statistical survey design for peer review, and amend where needed.

- Submit the proposal and review outcomes to the European Commission, as the surveys will be subject to DCF requirements.

- Implement the survey, archive and quality-assure the data, carry out the analysis of each national survey, document the results, and submit them for review (e.g. by ICES WGRFS) before provision to the ICES assessment EG.

- Obtain feedback from end users after the data have been used for assessments or other purposes, and amend future surveys (e.g. sampling intensity by stratum; collection of additional data) where needed to better address end user needs without compromising integrity of time series and subject to cost-benefit analysis.

- Advice is needed on post-release mortality refer to WGRFS but studies may be needed.

Table 25. Recreational fishery catches of seabass in ICES Areas IVb,c and VIIa,d-h, as reported to ICES.

(a) France		Kept	RSE	Released	RSE	Total	RSE	Release rate
2009-2011	NE Atlantic	2,343t		830t		3,173t	26%	26%
	ICES IV & VII	940t		332t		1,272t	>26%	26%
2011-2012	NE Atlantic	3,146t		776t		3,922t		20%

RSE was 26% for area VII and VIII combined; area VII represented 40% of total.

~ 80% by weight in 2009/11 was recreational sea angling

(b) Netherlands			Kept	RSE	Released	RSE	Total	RSE	Release rate
March 2010-Feb 2011	Southern North Sea	By number	227000	38%	127000	27%	354000	26%	64%
		By weight	138t	37%					
March 2012-Feb 2013	Southern North Sea	By number	335000	26%	332000	21%	667000	17%	50%
		By weight	229t	26%					

98% by weight is recreational sea angling.

(c) England		Kept	RSE	Released	RSE	Total	RSE	Release rate
2012	ICES IVbc, VIIa,d,e,f	230– 440t		150-250t		380 – 690t	26-38%	36-39%

Survey covered only recreational sea angling

Range of values is for different effort estimation procedures

4 Regional coordination of small scale and recreational fisheries data collection

Small scale and recreational fishing takes place mainly close to the coast, although some larger small scale, private and charter vessels (usually in the 10-12m category) are capable of operating quite far offshore. This coastal distribution means that fishing may impact some stocks that are largely inshore and of national interest, as for example coastal cod, mullets, sea breams and wrasse. At the same time, the fishery may harvest stocks or components of stocks that extend into international waters and are also harvested by large scale fleets from several countries (i.e. cod, whiting, flatfish, sea bass, and elasmobranchs). In some cases these are a minor part of the total international harvest, but in other cases their contribution to total catch is substantial. For example, recreational harvests of area IV&VII sea bass, western Baltic Cod and Norwegian coastal cod are around a quarter to a third of the total fishery harvest.

SSF are characterized by using a variety of gears, and by changing the fish targeted during the year. Fishing gears used by this fleet are mostly, but not exclusively, passive gears. In some countries, RFS are allowed to use commercial gears such as longlines and gillnets. This means that in some cases, the impact of the SSF and RFS in the by-catch of some PETS may be significant.

RCGs will be the responsible body to define regional sampling objectives. To this aim, end user's needs (estimates, aggregation and precision levels), the feasibility to collect such information and their potential impact on data collection programmes (available resources) will be taken into account. Once the regional objectives are defined, RCG will also need to determine the sample size and the optimal sampling effort allocation. WKPRECISE (ICES 2010b) and WKPICS2 (ICES 2013b) provided some general principles which should be taken into account also for SSF and RFS.

The regional approach to a sampling design proposed by fishPi involves sampling allocation to different countries to be optimised taking into account the national share of catches from international stocks. This will result in national coastal stocks being sampled primarily by the country in whose waters the stocks occur. However, it must be noted that this philosophy was originally proposed for the sampling of large scale fisheries, and that several issues must be taken into account when extending it to SSF and RFS:

4.1 Conflicts between national and regional data needs

Specific national measures may apply to waters within 6 nautical miles such as spatial closures or other targeted measures that apply mainly to the national commercial fleets or to recreational fishing. The management of national fisheries may therefore present specific sampling needs that are not required for management at the regional scale, or that may provide only small improvements to precision for shared international stocks.

Additionally, it should be noted that a regional sampling programme that is optimised to provide the required precision for the main international stocks throughout their geographic range, may provide inadequate data for national inshore management needs.

It is necessary therefore to consider the ability of national laboratories to accommodate a flexible approach to optimizing regional sampling allocations across countries for large scale fisheries whilst still maintaining the sampling programmes needed for inshore fishery management including small scale fisheries and recreational fisheries. If there are important SSF, as well as RFS, this will require substantial additional investment in surveys and there is potential to consider if there are common survey methods that could be applied to SSF and recreational fishing vessels to reduce costs.

This implies that in regions with significant SSF and RFS operating in inshore waters and subject to national inshore management schemes, regional coordination of sampling would have two stages:

- Firstly an evaluation of how sampling effort would best be allocated across countries and laboratories to deliver the data quality needed for shared international stocks subject to assessments;
- Secondly an evaluation of specific national requirements for data to support national inshore management schemes (which may also involve fisheries for shared international stocks) and to support assessment of inshore stocks fished only or predominantly by that country.

Given the availability of human and financial resources for fishery sampling in each country, there will be a need for a trade-off in sampling effort allocation between large scale and small scale commercial fleets and also recreational fisheries where data are required. There will never be a single optimised regional sampling programme, only a set of scenarios which need to be evaluated. To do this objectively, the types of analysis developed within fishPi and applied in the Case Studies in WP2, and possibly additional approaches such as management strategy evaluations, are essential to show the impact of altering the national allocation of commercial and recreational fishery sampling on the quality of estimates for each species and stock for which data are needed for assessment and management.

4.2 The need for reliable catch and effort estimates based on sampling

The optimization of sampling effort allocation among different countries, taking into account the national share of catches of international stocks, requires reliable estimates of small scale and recreational harvests of all the stocks to allow the relative catches of stocks by country and fishing method to be established.

When census data are not available, the estimation of the harvest for RFS and SSF requires methodological approaches that are very different from that of commercial fisheries. Surveys designs for these fisheries depend on what data are already available, such as the existence of vessel registers or lists of recreational license holders. They often involve two stages: firstly, a survey to describe and quantify the total population of vessels or fishers; and secondly, selection of a representative sample of the population to collect more detailed data on catches, effort, gears used and other information needed. The complexity of this process has to be taken into account.

The general concept of regional coordination of sampling proposed by fishPi could however apply where small scale and recreational harvests of assessed stocks in a region are known. If they are a very small part of the total fishery harvests for that species, it implies that catches could be monitored at relatively low precision without impacting the accuracy of the scientific stock assessments. In contrast, if small scale and/or recreational fisheries take a relatively large part of the catch, more precise estimates are needed. Regional Coordination Groups in collaboration with experts on sampling surveys (e.g. ICES WGRFS, ICES WGCATCH), will be able to advise on the need for SSF and/or RFS surveys and their degree of precision, depending on the available information on commercial and recreational harvests of all species in a region. To do this objectively, would require using procedures analogous to those examined in the fishPi project for sampling of commercial fleets, and simulation studies such as management strategy evaluations (MSE).

Currently, recreational catch estimates have been supplied for relatively few species in European waters, partly due to the very limited species list for such surveys in the current DCF Commission Decision. During 2015 ICES provided advice to the European Commission on the need for recreational fishery surveys, following a formal request, and advised that data were needed for all species so that the full range of recreational fishery impacts can be evaluated. In most cases, recreational fishery surveys can collect multispecies data with relatively little extra effort. Until such data are available, the concept of a regional coordination and optimization of recreational fishery surveys will be difficult to put in practice.

4.3 Cost benefit analysis

As the costs of sampling large scale and small scale commercial fisheries and recreational fisheries can differ widely, there is a need for an objective framework for evaluating the relationships between sampling effort (and costs) and precision achieved for each data set, and the relationship between precision of each estimate and the precision of stock assessment estimates using the data.

This is needed so that resources can be directed most efficiently to data collections that have greatest leverage on the quality of fishery management advice. ICES is currently hosting a workshop on cost benefit analysis of data collection in support of stock assessment and fishery management (WKCOSTBEN), the aim of which is to develop a cost-benefit framework that could, for example, help RCGs in developing advice on how sampling resources within a region could best be allocated between different types of data collection.

5 Interactions between small scale and recreational fisheries, and legislation other than cfp

5.1 Water Framework Directive (WFD)

The European Water Framework Directive (WFD) commits MSs to achieving good qualitative and quantitative status of all water bodies (including marine waters up to one nautical mile from shore) by 2015. The WFD includes River Basin Districts covering entire river catchments as natural transboundary units. For each waterbody type biological quality elements (i.e. macrophytes, fish, and invertebrates) are given and these should be included in MS monitoring programmes in order to determine the ‘good ecological status’ of specific waterbodies. Assessment of quality must be monitored through extensive national networks/programmes and must be evaluated against water type specific reference values. If good status cannot be achieved measures must then be taken to improve status toward good ecological status/conditions. The WFD should have been implemented in all MS by 2015. However although it varies between countries, its implementation is being delayed.

Marine recreational fisheries can have an impact on the WFD objectives if recreational activities impact on the Biological Quality Elements (BQE) in particular fish. However this is only the case in transitional waters where fish is currently included as a BQE. Hence changes in fish communities in coastal waters are currently not taken into consideration when evaluating ecological status or only indirectly if recreational fishing result in negative consequences for other BQEs. In the WFD commercial fisheries (including large scale and small scale) are considered a potential “pressure” meaning that this is considered a potential threat to good ecological status. However, at present this seems not to be the case for recreational fisheries, although not clearly described in the WFD text and related annexes. The aim of the WFD is to protect and improve the ecological status of water bodies and for many fish species potentially also improving the recreational fishery. Stocking or unintentional introductions of non-indigenous species is also considered a pressure under WFD, and a successful implementation of WFD could reduce the occurrence of such stockings and introductions.

5.2 Marine Strategy Framework Directive (MSFD)

The European Marine Strategy Framework Directive (MSFD) was developed to provide a framework for MS to protect the marine environment more effectively (EU, 2008). This is to be done by maintaining biodiversity and providing diverse and dynamic oceans, which are clean and healthy, while allowing the sustainable use of marine resources (EU, 2008). The MSFD is based on an ecosystem approach and will, where necessary and appropriate, draw on existing regulation in order to achieve coherence between policy areas (e.g. CFP EU, 2013; Habitats Directive CEC, 1992). It came into force in 2008, and aims to allow MSs to take the necessary measures to achieve or maintain Good Environmental Status (GES) by 2020. European marine regions were defined for the purpose of monitoring water status and developing actions to achieve GES (e.g. NE Atlantic Ocean, Mediterranean Sea, Black Sea, Baltic Sea), with sub-regions also defined in the Northeast Atlantic and Mediterranean. In order to meet the requirements of the Directive, MS are obliged to cooperate with others in the same (sub)region, including through the relevant Regional Sea Conventions. Cooperation is also required between MSFD regions in order to ensure consistency and coherence across the EU.

To determine GES, eleven qualitative descriptors of the state of the environment were defined comprising biodiversity, non-indigenous species, commercial fish, foodwebs, eutrophication, seabed integrity, hydrographical conditions, contaminants, food safety, litter, and underwater noise (EU, 2008). The European Commission (EC) developed a detailed set of criteria and indicators to help interpret these descriptors (COM, 2010). There are there potential interactions between SSF/RSF and MSFD: positive benefits of legislation on SSF/RSF; positive benefits of SSF/RSF in delivering legislation; and impacts of SSF/RSF on delivery of legislation. Currently, there is no mention of RSF in either the MS assessments or the Programme of Measures (POMs), possibly due to the regional scale impacts considered within MSFD and the mitigation under existing legislation like CFP (EU, 2013). The most likely interactions between SSF/RSF and MSFD are under D1 biodiversity, D3 fish stocks in relation to fishing mortality and population structure, D4 food webs in relation to uncertainty and anthropogenic pressures; and D6 seabed integrity for passive fishing gears. Workshops are being planned by ICES in 2016 that provide an opportunity to engage with the process, and should be attended by experts in RSF. SSF are included inside the term commercial fisheries, but attention should be made to the differences between large scale and small scale fisheries and their different impacts.

There is also the potential for time-series of recreational fishing data to be used in the development of indicators of GES, so it is worth considering how monitoring of recreational fishing may provide evidence to underpin indicators or monitor progress against indicators in future. For example, simple additional information could provide data that will underpin both DCF and MSFD monitoring needs. It is also important to develop thinking around how SSF/RSF should engage with future MSFD cycles.

5.3 Maritime Spatial Planning (MSP)

Maritime spatial planning (MSP) has been succinctly defined as the "Public process of analysing and allocating spatial and temporal distribution of human activities in marine areas to achieve ecological, economic and social objectives […] through a political process” (Ehler and Douvere, 2007). The European Union policy instrument for MSP is Directive 2014/89/EU which gives the definition as “A framework under the Integrated Maritime Policy to manage human activities at sea & in coastal zones” (EC, 2014). Other relevant policy instruments include The United Nations Convention on the Law of the Sea (UNCLOS 2014), the MSFD (2008/56/EC), the Habitats Directive (92/43/EEC) and Birds Directive (2009/147/EC). Marine related legal instruments such as the MSFD are likely consumers of MSP outputs and do not explicitly legislate on MSP.

MSP is believed to reduce conflicts, encourage investments, increase coordination between administrations, increase cross-border cooperation between EU countries and protect the marine environment. The financial provision for MS to implement Directive 2014/89/EU is delivered by European Structural and Investment Funds, including the European Maritime Fisheries Fund (EMFF). No explicit reference to small scale (SSF) and recreational sea fishing (RSF) is made by MSP (2014/89/EU), but their consideration is implicit under the tourism and fisheries sectors. Their relevance for MSP is of particular relevance in the coastal area. Their reduced mobility makes them extremely dependent on coastal ecosystem resources, and conflicts can arise between SSF, large scale and recreational fishing fleets (and also with other uses of the sea such as renewable energies, dredging, tourism, and even conservation marine areas), because they interact in the same area. The relevant minimum requirements for SSF and RSF within MSP are covered under:

- Article 6 takes into account environmental, economic and social aspects, ensures the involvement of stakeholders, and makes the use of the best available data.
- Article 8 establishes national maritime spatial plans that must account for relevant interactions between activities and across uses.
- Article 9 enshrines stakeholder engagement in decision-making.

Recent assessments of recreational catches and activity levels in specific fisheries evidence the importance of including RSF in MSP under the above listed articles and planning should also consider incidental and non-fishing recreational fisher activity (e.g. bait collection and foot-fall) with any associated ecosystem effects where both relevant and practicable. Despite RSF being frequently perceived as a predominantly extractive activity, current evidence indicates that the sector is of socio-economic importance and that the ecological impact might be significant. Recreational fishing has been linked to positive health and wellbeing in participants and can generate economic benefits through direct and secondary effects. Hence the sympathetic management of resources valued by recreational fishers will contribute to the fulfillment of the MSP.

MSP planners require digitized maps of spatial and temporal SSF and RSF effort. Maps should be of sufficient temporal and spatial resolution to enable informed decision-making in consideration of coastal and marine habitats and the activity of other marine stakeholders. Other potentially relevant factors should be recorded with spatial and temporal activity data, for example fishing platform (e.g. for-hire or shore sectors) and key aspirational and caught marine species. In making value judgments on use policy, economic maps are critical, particularly where economically significant activity is thought likely to occur. SSF and RSF stakeholders need to be engaged to understand what and where they value within their marine environment (e.g. fishing areas, fishing types, target species). There is a general need to understand what conflicts and compatibilities exist, or could arise, with other marine activities. In addition, stakeholders will have valuable insights on how to maximize the socio-economic value derived from the recreational sea fishing sector.

All MS require provision for a maritime spatial plan covering all marine activities. The objective is to balance stakeholder interests, economics and environmental management of marine resources. Maritime Spatial Planning and associated recommendations and legislation are evolving, but a description of the activities and economics of small scale and recreational sea fisheries in time and space will be critical to the maritime spatial planning process. It is important (as specified in the EU directives) for the process to be transparent and stakeholder driven.

6 Conclusions

Regional coordination of small scale and recreational fishery data collection programmes is needed to ensure that end users are supplied with the catch estimates or other data needed by them, at the required spatial resolution, temporal coverage, and quality. Coordination is a role for the lead scientists for the surveys in each country, the Regional Coordination Groups, the ICES Working Group on Recreational Fisheries Surveys and the ICES Working Group on Commercial Catches (as technical expert advisory groups).

For SSF, data collected under the Control Regulation via logbooks and sales notes are limited, and sampling is needed to improve the quality of collected information. For RFS, data on catches and effort are always dependent on sampling. The appropriate design of RFS and SSF sampling programmes in a region do not necessarily have to be completely harmonised between countries. The most important attribute is that surveys have robust statistical designs to minimize bias and allow correct calculation of precision. Catch estimates from different surveys can then be combined. Information must be collected to allow evaluation of the potential for bias, for example non-response rates, refusals and incomplete coverage. The types of data needed from SSF and RFS should be defined by end-user needs to achieve Common Fisheries Policy goals as well as to support national inshore management.

RFS and SSF surveys can be complex and are statistically demanding. With multiple countries contributing to these estimates, a Quality Assurance Framework (QAF) for documenting and archiving data quality is required at a national and regional level. Regional survey programmes should be subject to expert peer review before implementation and amended where necessary. The results of surveys should also be subject to review, for example by ICES WGRFS, ICES WGCATCH, or by other independent experts.

The methods proposed by fishPi to explore how sampling of commercial fisheries could be optimised regionally by appropriate allocation of sampling effort between countries cannot be used directly to allocate sampling efforts between large scale, small scale and recreational fishery surveys. This is due to the differences in sampling methods and to the very nature of the fisheries.

Firstly, it should be noted that a regional sampling programme for the main international stocks throughout their geographic range, may provide inadequate data for national inshore management needs for SSF and RFS. The approach should therefore be flexible enough for national laboratories to accommodate regional sampling effort allocation for large scale fisheries whilst still maintaining the sampling programmes needed for inshore fishery management.

Secondly, the optimization of sampling allocation among different countries, taking into account the national share of catches of international stocks, requires reliable estimates of small scale and recreational harvests of all the stocks. When census data are not available, the estimation of the harvest for RFS and SSF requires complex methodological approaches that are very different from that of commercial fisheries (They often involve two stages: firstly, a survey to describe and quantify the total population of vessels or fishers; and secondly, selection of a representative sample of the population to collect more detailed data on catches, effort, etc).

Despite these considerations, the general concept of regional coordination of sampling proposed by fishPi could be applied. Depending on the information available, Regional Coordination Groups (together with WGCATCH and WGRFS) would be able to advice on the need for SSF and/or RFS surveys and their degree of precision. Surveys at relatively high precision will be needed if SSF or RFS are a main source of fishing mortality, and monitoring at lower precision will be sufficient if they are not. In this sense, the tasks of RCGs would be facilitated if they have access to recreational catch estimates for a much wider range of species in a region than at present, so that the full impact in all the fisheries can be evaluated.

In relation with a regional data base, a distinction should be made between transversal variables collected through a census approach (Control Regulation or ad-hoc methods for SSF), and those transversal variables collected through sampling programmes designed for RFS and SSF. Whereas the first group apparently can be adapted to the current structure of the RDB (FishFrame), the second group needs fundamental changes in the structure of the data base. The potential for a diverse range of off-site and on-site survey methods in a region means that the concept of a regional data base for recreational and small scale fishery data may not be easily achievable. Further work will be needed to determine the needs and the design of such a data base. At this stage, a principal focus should be to ensure that data from national surveys of different types are properly archived and subject to appropriate QA/QC procedures, and that the full process from survey design, implementation, data archiving and quality control, data analysis and reporting is fully documented and transparent for each country contributing to a regionally coordinated recreational survey programme.

References

Anon 2013. Report of the Working group on Common understanding and statistical methodologies to estimate/re-evaluate transversal data in small-scale fisheries. 21-23 May 2013, Nantes, France. 76pp

Anon 2014. 3rd Planning Group on Economic Issues (PGECON). 31 Mar – 4 Apr 2014. Berlin, Germany. 41 pp.

Armstrong, M., Brown, A., Hargreaves, J., Hyder, K., Pilgrim-Morrison, S., Munday, M., Proctor, S., Roberts, A., Williamson, K. 2013. Sea Angling 2012 – a survey of recreational sea angling activity and economic value in England. Synthesis report. http://www.marinemanagement.org.uk/seaangling/

CEC. 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Council of the European Communities (CEC). Official Journal of the European Communities, 206, 7–50.

COM. 2010. Commission Decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters (2010/477/EU). European Commission (COM). Official Journal of the European Union, 232(2008), 14–24. Retrieved fromhttp://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:232:0014:0024:EN:PDF.

Connelly, N. A., and Brown, T. L. 1995. Use of angler diaries to examine biases associated with 12-month recall on mail questionnaires. Transactions of the American Fisheries Society, 124: 314–422.

Demanèche, S. 2013. Methodological issues to estimate catches and fishing effort of smallscale fisheries by sampling fishing trips on-site. 7th International Fisheries Observer & Monitoring, 8–12 April 2013, Vina Del Mar (Chile), http://www.ifomc.com/panelists4.html

Ehler, C., Douvere, F. 2007. Visions for a sea change. Report of the first international workshop on marine spatial planning. Intergovernmental Oceanographic Commission and the Man and the Biosphere Programme, UNESCO. Paris, France.

EU. 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Official Journal of the European Union, 164, 19–40.

EU. 2013. Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC. Official Journal of the European Union, 354, 22–61.

EU. 2014. Directive 2014/89/EU of the European Parliament and of the Council of 23 July 2014 - establishing a framework for maritime spatial planning. Official Journal of the European Union.

García-Flórez, L., Morales, J., Gaspar, M.B., Castilla, D., Mugerza, E., Berthou, P., de la Fuente, L.G., Oliveira, M., Moreno, O., del Hoyo, J.J.G. and Arregi, L., 2014. A novel and simple approach to define artisanal fisheries in Europe. Marine Policy, 44, pp.152-159.

Guyader, O., Berthou, P., Koutsikopoulos, C., Alban, F., Demaneche, S., Gaspar, M.B., Eschbaum, R., Fahy, E., Tully, O., Reynal, L. and Curtil, O., 2013. Small scale fisheries in Europe: A comparative analysis based on a selection of case studies. Fisheries Research, 140, pp.1-13.

ICES 2010a. Report on the Planning Group on recreational Fisheries (PGRFS), 7-11 June 2010, Bergen, Norway. ICES CM 2010/ACOM:34. 168 pp.

ICES 2010b. Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE), 8-11 September 2009, Copenhagen, Denmark. ICES CM 2009/ACOM:40. 43 pp.

ICES 2011. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS) , 27 June - 1 July 2011, ICES Headquarters, Denmark. ICES CM 2011/ACOM: 50. 116 pp.

ICES 2012a. Report of the Study Group on Practical Implementation on Discard Sampling Plans (SGPIDS), 18–22 June 2012, ICES HQ, Copenhagen, Denmark. ICES CM 2012/ACOM:51. 87 pp.

ICES 2012b. Report of the Working Group on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS), 8 - 10 November 2011, Bilbao, Spain. ICES CM 2011 / ACOM:52. 55 pp.

ICES 2012c. Report of the Working Group on Recreational Fisheries Surveys (WGRFS), 7 – 11 May 2012, Esporles, Spain . ICES CM 2012 / ACOM:23. 55 pp.

ICES 2013a. Report of the ICES Working Group on Recreational Fisheries Surveys (WGRFS), 22-26 April 2013, Esporles, Spain. ICES CM 2013/ACOM:23. 49 pp.

ICES 2013b. Report of the second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS2), 6 - 9 November 2012, ICES Copenhagen. ICES CM 2012 / ACOM:54 71 pp.

ICES 2014a. Report of the Inter-Benchmark Protocol for Sea Bass in the Irish Sea, Celtic Sea, English Channel, and Southern North Sea (IBP–Bass). By correspondence. ICES CM 2014/ ACOM:46.

ICES 2014b. Report of the Working Group for the Celtic Seas Ecoregion (WGCSE), 13–22 May 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:12.

ICES 2015a. Report of the Working Group on Commercial Catches (WGCATCH), 10-
14 November 2014, ICES HQ, Copenhagen, Denmark. ICES CM 2014\ACOM:36. 295 pp.

ICES 2015b. Report of the Working Group on Recreational Fisheries Surveys (WGRFS), 1–5 June 2015, Sukarrieta, Spain. ICES CM 2015\SSGIEOM:10. 111 pp.

ICES 2016. Report of the Working Group on Commercial Catches (WGCATCH), 9-13. November 2015, Lisbon, Portugal. ICES CM 2015/SSGIEOM:34. 111 pp.

JRC 2016. Report on the 2nd Workshop on Transversal Variables. 22-26 February 2016. Nicosia, Cyprus. 108pp

Lyle, J., Coleman, A. P. M., West, L., Campbell, D., and Henry, G. W. 2002. New large-scaled survey methods for evaluating sport fisheries. In Recreational Fisheries: Ecological, Economic and Social Evaluation, pp. 207–226. Ed. by T. J. Pitcher, and C. Hollingworth. Blackwell Publishing, Oxford. 288 pp.

Pawson, M.G., Tingley, D., Padda, G. and Glenn. H. 2007. Final Report: EU contract FISH/2004/011 on Sport Fisheries (or Marine Recreational Fisheries) in the EU (European Commission Directorate-General for Fisheries)

Pollock, K. H., Jones, C. M., and Brown, T. L. 1994. Angler survey methods and their aplications in fisheries management. American Fisheries Society, Special Publication 25, Bethesda, MD.

Tarrant, M. A., and Manfredo, M. J. 1993. Digit preference, recall bias and nonresponse bias in self reports of angling participation. Leisure Sciences, 15: 231–238.

Van der Hammen, T and de Graaf, M. 2012. Recreational fishery in the Netherlands: demographics and catch estimates in marine and freshwater. IMARES Wageningen UR, Report Number C147/13.

Vaske, J., Huan, T. C., and Beaman, J. 2003. The use of multiples in anglers' recall of participation and harvest estimates: some results and implications. Leisure Sciences, 25: 399–409.

Vølstad, J. H., Afonso, P. S., Baloi, A. P., de Premegi, N., Meisfjord, J., & Cardinale, M. 2014.Probability-based survey to monitor catch and effort in coastal small-scale fisheries. Fisheries Research, 151, 39–46.

Appendix 1 WP3.3 Small scale fisheries and recreational fisheries survey form

Year	2012
Member State:	
Institute:	
Name:	

Note to MS that have several institutes covering different regions/countries (e.g., UK, Spain): please coordinate your answers to this table in order to avoid duplication of fleets and other values.

	Values			Data source		
Vessel length classes	<10 m	10-12 m	>=12 m	<10 m	10-12 m	>=12 m
No. of vessels (EU register) (*)				EU register	EU register	EU register
No. of licensed vessels						
No. of actively fishing vessels						
No. of days at sea						
No. of trips						
Fish landings in weight (tons)						
Shellfish landings in weight (tons)						
Fish landings in value (euros*1000)						
Shellfish landings in value (euros*1000)						

(*) To obtain this information, please use the following link:
http://ec.europa.eu/fisheries/fleet/index.cfm?method=Download.menu
and plug in your country code, pressing "Active at a date" and setting it to 01/07/2012.

Year:	Present sampling programme
Member State:	
Institute:	
Name:	

SSF = Small Scale Fisheries

LSF = Large Scale Fisheries

	catch composition	length frequency	age	maturity	spatial mapping
0. Does your sampling programme consider SSF as the vessels <12 m? If not, please detail the definition you use.					
1. Are SSF sampled? If yes, detail the year that sampling started.					
2.0 If yes, are SSF sampled randomly as part of a general sampling programme for commercial fisheries, or through a dedicated sampling programme?					
2.1 Provide a brief description of the sampling of SSF under that programme					
2.2 Are you aware of studies in your MS that evidenced differences in the variable between SSF and LSF? If yes, supply references (and upload document to sharepoint)					
3.0 Provide a description of potential biases involved in your sampling of the SSF for each variable					
3.1 Provide a qualitative assessment about data consistency and biases involved in sampling the SSF for each variable					

ANNEX 18 WP4 Guidelines and calendar for the application of quality indicator functions

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP4 – Data quality at a national and regional level

Deliverable 4.1 - Guidelines for the application of quality indicator functions and routines applicable on national and regional datasets, and Calendar for applying the Quality Checks at a national and regional level.

Lead: IFREMER, France

Core Team: Joël Vigneau, Laurent Dubroca, Kelig Mahé (IFREMER, France)
Alastair Pout, Liz Clarke (MSS, United Kingdom)
Chun Chen (IMARES, Netherlands)
Mike Armstrong, Ana Ribeiro Santos (CEFAS, UK)
Maria Hansson (SLU, Sweden)
Eva Velasco (IEO, Spain)
Christoph Stransky (Thunen, Germany)

Table of Contents

Executive summary

The objectives of WP4 were to develop guidelines to evaluate the quality of data at national and regional levels using shared tools and to agree on timetables for the implementation on annual quality checks. The work in this WP was conducted together with the WP4 of the same project run in the Mediterranean and Black Sea, as agreed at the kick-off meeting with the EU. The data structure (called csPi data structure) upon which the work was developed was an upgrade of the COST/FishFrame format (ICES, 2008), as agreed in the SC-RDB (ICES, 2014) and further developed in WP2. The source code of all the functions is available on a free access website (https://github.com/ldbk/fishPifct), and it is possible to compile a library directly from the website.
The Guidelines on the use of the quality checks functions were structured as follows:

- Verification of the data structure. Import csv files into csPi structure with integrity control of the tables, and check every field against reference tables and agreed list of entries.
- Verification of the consistency of the information populated into the database, e.g. identification of trips without fishing operations, missing sampled weight.
- Advanced data checks. Smart outlier detection of outliers in the numerical fields, identification of errors in fishing operation positioning
- Exploratory data analysis. Production of maps, figures and reports, adapted from the COST library exploratory analyses functions.

These functions and the library are to be considered as the first version of a library which aims to improve with time and serves the needs of any end-user at national or regional levels. Moreover, the data structure upon which the functions were developed is also likely to evolve in the near future, so the functions were developed in such a way to accommodate and let regional coordination groups (RCG) the liberty to agree on any structure, field names, field boundaries or reference lists for any component of the data structure. Identically, the reporting procedure of the quality checks adds a column to each of the tables of the data structure and indicates the issues at the record level, when relevant.

A calendar for applying the Quality Control at a national and regional level was proposed, taking account of the end-user needs and the national constraints. The ideal time frame would require Member States to run the quality controls at their level at each quarter and validate all their datasets by the end of February. The first half of March would then be used to populate the regional database and run the quality checks at the regional level. Adaptation of this time frame will have to be considered, since field constraints make it impossible for all Member States to be in a position to transmit validated data for all of their datasets by the end of February.

1 Background and Context

The objectives of Work package 4 were to evaluate the quality of data collected at national and regional levels using shared tools and progress on the harmonisation of data exchange formats and data structures. Another objective was to agree timetables for the implementation of annual quality checks. The data structure upon which the work was developed is an upgrade of the COST/FishFrame format (ICES, 2008), as agreed in the SC-RDB (ICES, 2014) and further developed in fishPi WP2 recently (see also WP2 deliverable).

The fishPi kick-off meeting took place in April 2015 together with the same project in the Mediterranean and the two projects were asked to share their work on the implementation of quality check functions.. Two WP4 members of the Mediterranean project participated to the WP meeting in Port-en-Bessin (July 2015, France) and joined the fishPi team for development of the functions. The minutes of the meeting was used for reporting on WP4 for both projects, and the specific development planned for the Mediterranean will not be cited here.

Since the Port-en-Bessin meeting, the quality check functions have been developed and made available on the github website (https://github.com/ldbk/fishPifct). These functions and the library are to be considered as the first version of a library meant to improve as time goes on.

2 Guidelines and best practices for quality checks and data validation

2.1 Review of Guidelines and best practices from previous WG

The scope of WP4 included consulting the data quality procedures and best practice guidelines set out in the reports of various ICES expert groups noticeably WKPICS, SGPIDS, WGCATCH and related work undertaken in the RCMs, and compiling a list of data quality checks that can be used for national data checking and regional data checking, specifying a set of minimum data quality checks required on national data sets.

The Regional Coordination Meeting for the North Sea and Eastern Arctic (RCM (NS&EA), 2014) summarised the stages in data quality assurance and quality control for fishery sampling schemes, and who is responsible. It showed that the design, collection, quality control and use of Data Collection Framework data in Europe involves many stages and groups of people, and that in general the procedure should follow a well-defined series of steps which are described by the RCM. The report of the ICES Working Group on Commercial Catches (WGCATCH: 2014) subsequently provided an illustrative schema for possible elements of quality evaluation of a fishery sampling programme (Table 1). It follows the successive stages given by the RCM (NS&EA) for designing a sampling scheme, implementing it, archiving and extracting the data, and the analysis of the data to provide estimates for end users. This breakdown of a sampling scheme by design, implementation, data capture etc. has been central to recommendations by the ICES WKPICS, SGPIDS, PGCCDBS and (now) WGCATCH for procedures to provide quality indicators appropriate to each of the stages. These groups have provided guidelines for quality evaluation at each stage, and the sources for these are given below.

2.2 Design of the sampling scheme

Design principles for marine fishery sampling schemes in Europe that are described and promoted in the fishPi project (WP2) are covered in detail in the ICES Workshop on Methods for Merging Métiers for Fishery Based Sampling (WKMERGE: ICES 2010), the ICES Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS: ICES 2011a, 2012a, 2013a), the ICES Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS: ICES 2011b, 2012b, 2013b), and also the ICES Working group on Recreational Fishery Surveys (WGRFS: ICES 2012c, 2013c, 2014b). These ICES expert groups have promoted the adoption of sampling schemes that follow well-established statistical design principles, and avoiding ad-hoc sampling methods that are prone to bias that may be difficult to quantify and may vary with time. The groups have provided clear definitions of the population to be sampled, sampling frames for accessing the population, primary and lower-level sampling units, stratification of the sampling units, and sample selection procedures that allow control over sampling probabilities. ICES WKPICS (ICES, 2012a) and WGRFS (ICES, 2013c) used these design principles to draw up best practice guidelines for designing commercial and recreational fishery sampling surveys. Such guidelines were intended to help countries design sampling programmes, but could also be used for external auditing of the quality of sampling designs against best practice. The ICES WKPICS, SGPIDS and WGRFS report series provide detailed information on different possible designs of fishery sampling schemes on shore and at sea, and how these relate to the characteristics of the national fisheries. Practical considerations for designing schemes are considered by each group. The series of ICES SGPIDS reports also provide detailed recommendations on how sampling should be conducted at sea to ensure representative sampling within hauls within a trip. Practical implementation of shore-based sampling of commercial fisheries has been covered by the ICES WGCATCH (ICES 2014a).

More general guidelines on survey design covering the same design principles as proposed for fisheries sampling can be found on the Eurostat web site at:
http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-08-003

2.3 Implementation of sampling schemes

Even if the overall sampling design follows best practice, problems may emerge at the stage of implementation that can lead to unintended bias and reduced precision. This may happen if staff deviate from sampling protocols, or if factors such as refusal rates, inadequate resourcing or staff availability, or altered fishing behaviour when observers are on board, lead to non-representative sampling or problems with missing or inadequate numbers of observations for one or more strata. Providing indicators of data quality related to implementation error (which have bias and precision components), and which can be used by stock assessors or other end users, has proved challenging. The ICES Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU: ICES 2008) developed a detailed questionnaire to identify the potential for bias in all the stages of implementing a fishery sampling scheme. This is comprehensive and useful to help national laboratories eliminate such biases as far as possible, by good practice. However it has proved too detailed for end users such as stock assessors. The ICES Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS: ICES 2011c) initiated a process of developing data quality reports to provide higher-level quality indicators for fishery sampling data within years and across time series. The within-year data quality reports underwent successive development by SGPIDS, WKPICS and WGRFS but have yet to find a routine application within the ICES stock assessment process. The new ICES Planning Group on Data Needs for Assessments and Advice has now taken up the challenge of developing a framework for ICES benchmark stock assessment workshops to document and use information on data quality including how data quality has changed over time.

2.4 Data analysis

Design-based sampling schemes require methods of data analysis that correctly represent the sampling probabilities at each hierarchical stage in sampling, and properly account for cluster sampling which is characteristic of most fishery sampling schemes. ICES WKPICS (ICES 2012a, 2013a) identified four principal design classes for sampling of commercial fisheries at sea or on shore, and provided detailed guidelines on design and analysis of data from each class of sampling including use of auxiliary data for re-weighting. fishPi Work Package 2 provides detailed implementation of these types of schemes in a regional context.

Programme stage	Existing guidelines and standards ("best practice")	Quality evaluation procedure	Performance measures	Possible Quality Indicators
Design of sampling scheme	e.g. WKPICS and WGRFS best practice guidelines; IBTS protocols etc.	Review of documentation on sampling design relative to quality standards	Indicators of bias potential due to design.	Score against quality standards, e.g. frame coverage, sample selection procedures etc.
Implementation of sampling scheme	e.g. WKPICS and WGRFS best practice guidelines; IBTS protocols etc.	Review of sampling outcomes – e.g. diagnostics of coverage, refusal rates, sample numbers and precision etc.	Indicators of extent of bias (e.g. low, medium, high, unknown); Indicators of precision.	Number of primary sampling units sampled in each sampling stratum; CV; frame coverage; refusal rates.
Data archiving and extraction	e.g. RCM North Sea and Eastern Arctic 2014 lists of data checks.	Review of documentation of QA/QC procedures relative to quality standards. e.g. use of electronic data capture; error traps etc.	Indicators of extent and effectiveness of QA/QC procedures.	Score against quality standards
Data analysis	e.g. WKPICS and WGRFS best practice guidelines; IBTS protocols; etc.	Review of documentation of estimation procedures relative to quality standards.	Indicators of extent of bias (e.g. low, medium, high, unknown)	Score against quality standards, e.g. analysis follows design

Table 1. Possible elements of quality evaluation of a fishery or survey sampling programme (ICES (2014)

2.5 Proposal of a general process for a fully functional RDB

In the RCM NS&EA 2014 meeting, data quality was said to be an issue in all steps involved in data-collection and the workflow involving seven steps were defined as follows.

This is a relevant description of the workflow and of all components which have an impact on data quality. In the report it was further described what kind of quality issues, QA/QC procedure, example diagnostics and recommended checks involved for each step (RCM NS&EA 2014). In order to reach a regional coordinated sampling program producing high quality data, all steps need to be taken care of and for the data-collection community it is important to elaborate and develop procedures for every step, each involving different kinds of actions to be taken. However, the workflow above illustrates the ideal situation where the collation of data is conducted within the RDB, which is not the present situation. In the current situation, collation of data is done by each country in different, unchecked ways and raised data are uploaded to the ICES database Intercatch. Quality checks are made during the upload process and there are also some controls available at a second stage once the data is uploaded.

One of the main tasks for WP4 in fishPi project is to focus on establishing quality checks on national data, e.g. to establish a list of necessary and valuable checks and to produce R code for making these checks. This work fits into steps 3 and 4 and to some extent step 5 above (National data-capture, National data processing and upload to RDB). More detail regarding the suggested checks is specified in section 4.

Quality checks already being develop within WebGr (checking quality of biological parameters) should be a part of the quality checks to be performed within step 3. The fishPi project strongly recommend that the WebGr tool is maintained and developed further since the tool is of great value to serve the quality checking procedure.

Guidelines for setting up proper sampling design and the documentation of it (Step 1 and 2) will be elaborated on in WP2 and within the case studies.

In order to get an overview on the data quality and the procedures used within a region, there is a need to establish structured documentation, such as Quality Reports. The outcome of the fishPi project will give valuable input in such a report but there is a need for RCM /RCG involvement to establish such documents. The "Eurostat Standard for Statistics Handbook for Quality Reports" should be used as a reference for such work.

2.6 WebGR services

WebGR (Web services for support of Growth and Reproduction Studies; http://webgr.azti.es/ce/search/myce) is Open Source software developed in 2008 by a consortium of research institutes and software developers from Federal Agency for Agriculture and Food (Germany), which has been developed during only one European project (DG MARE tender FISH/2007/07 Lot 1). This tool has the goal to be used for age calibration workshops and otolith exchanges as well as maturity staging. WebGR has the advantage that it can be used in similar fashion to Paint Shop Pro and GIMP, but instead of creating a layer for each reader in a specified file format, WebGR saves each reader's annotation of each image as a set of xy-coordinates that can be mapped on to that image, but the original image and the associated metadata remain unaltered. The use of WebGR will make workshops and exchanges more efficient and economic. Furthermore, it can contain reference collections with agreed ages, which can be used by inexperienced readers as a self-training tool, where they can access images and compare their annotation of images with those of experts. The programme has been used for several workshops and exchanges since 2010.

The coordinator of an age calibration workshop uploads the selected images to the server and it stores the images and metadata grouped by species, date, area etc. All participants of that workshop annotate and assign an age to each individual image without having access to the work of the other participants. When all the images have been annotated and aged, the coordinator arranges access for the participants to all the annotations and aged images. The participants can then compare and discuss each other's annotations and ages to identify sources of disagreement. However, the use of the WebGR tool for the exchange present some limits because it is a result of a single European project without update since the end of this project:

- WebGR is not a very intuitive tool
- WebGR could be jammed and there have been some crashes during the last years (WebGR is at the moment hosted by AZTI (Spain), who provide necessary help on the use of the programme voluntarily and in their own free time. However, AZTI personnel have not been able to make improvements to the programme since 2010).
- Only the participants of the exchange can view the images
- Only the exchange coordinator can extract individual statistical outputs (Average Percent Error, CV, SD and Variance).
- All exchanges and workshops using WebGR were not validated by the expert group WGBIOP (Working Group on Biological Parameters) as usable for Quality Control

As ICES is recommending the use of WebGR, there is a strong need for both development and improvement. One of the first developments would be a public access to summary statistics and controlled access to anonymised and peer-reviewed data.

A reference collection of otoliths and maturity reading would also be needed in order to help all sampling teams. These reference collections are not part of the current WebGr, this issue should be addressed by the future RDGs linking with ICES/WGBIOP.

3 csPi dataset

3.1 Structure and contents

The data structure (Fig. 1) used for developing functions in WP4 was specified during the WK-RDB 2014 (ICES, 2014c) and is an extension of the structure developed during the COST project (Anon, 2009) with additional fields to add information about the sampling strategies used.
Those additions are mainly :

- A new « Sampling Event » table (SE) which aims to provide a location for recording information relating to the provenance of the sample, the sampling design and the primary sampling unit.
- In all the tables, fields are added for a unique primary sampling unit identifier for the sample, relating to the new SE table.
- With several years of feedback on the COST format, some fields were added to handle more information and give more options for statistical estimates.

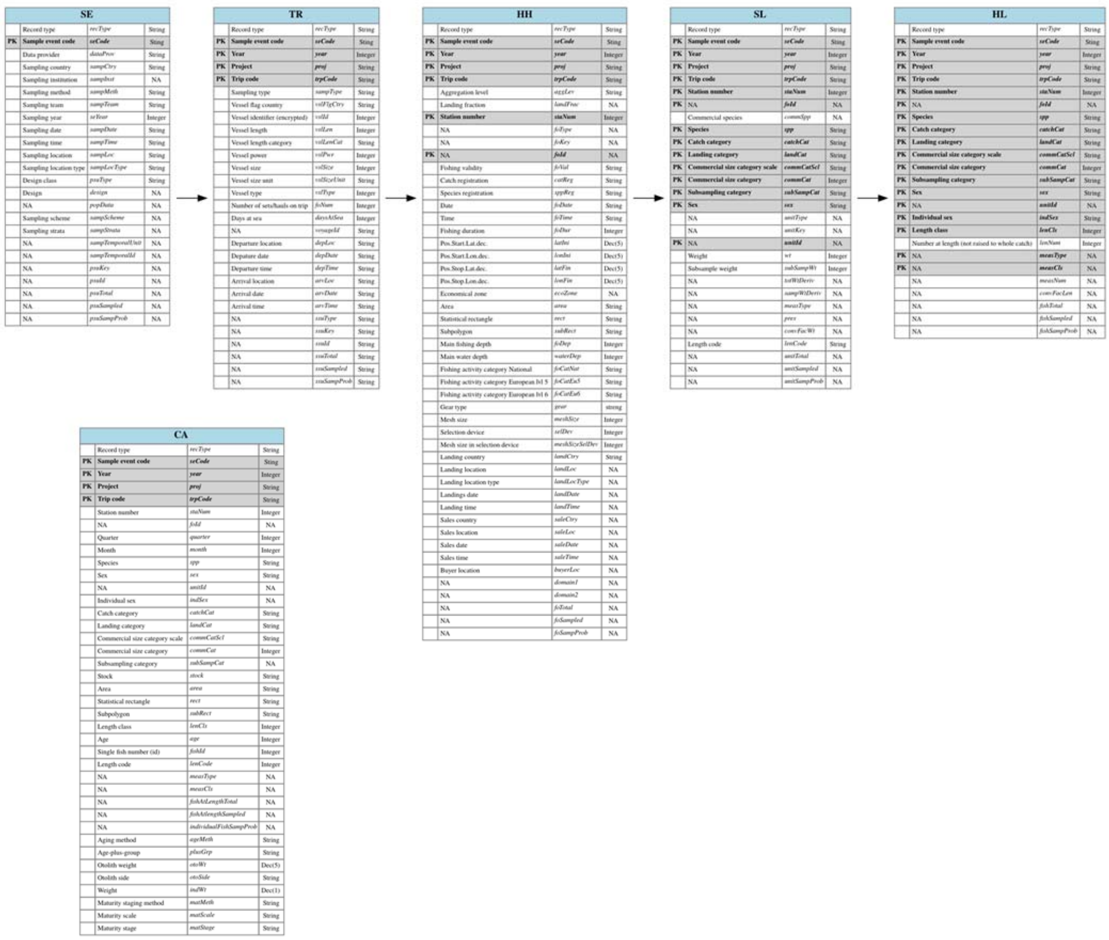


Figure 1. Data structure used for developing functions in WP4.

3.2 Example data sets

The « sole » COST example dataset was updated to the new data structure, those data are about sole fished by France in 2007 on ICES areas « 27.7.d » and « 27.7.e ».

Also the selected test cases for the fishPi project and the Mediterranean project were used as « testing » data sets (more detail provided in each case study report):

- For the Mediterranean project:
    - hake and red mullet stocks in the Gulf of Lions (GSA07)
    - small pelagics in GSA17

- For the fishPi project WP4 :
    - Case Study 1 Small Pelagic fisheries, vessels using pelagic fishing methods. (i.e. single and twin midwater trawls and pelagic seines; gears OTM, PTM, PS), all ICES area in FAO area 27, target species: herring, mackerel and sprat.
    - Case Study 2 North Sea Demersal, vessels using demersal fishing techniques (gillnets, hooks & lines, demersal trawls and demersal seines, beam trawls, i.e. all gears defined in vessel type (see below) with the exception of DRB FPOand TM regardless of the métier defined target assemblage, areas are ICES divisions IIIa, IV, VIa, VIb & VIId and the target species: all species of fish and Nephrops, but no other shellfish.
    - Case Study 3 Southern North Sea flatfish, vessels using demersal fishing techniques (gillnets, hooks & lines, demersal trawls and demersal seines, beam trawls), areas are ICES divisions IIIa, IV, & VIId and target species: plaice, sole, turbot & brill.
    - Case Study 4 Hake, vessels using demersal fishing techniques (mainly trawls, set gillnets and set longlines), areas are Division IIIa, ,Subareas IV, VI, VII and VIII, and Division Ixa and target species: hake

3.3 Quality checks framework

Fishery data have to be inspected in order to detect errors before using them in the stock assessment procedure (Chen 2003). Finding and correcting errors is one of the first tasks one needs to perform on a dataset in this case. Currently these checks are made at national level using mainly manual methods (based on graphs and numerical summary, see Vigneau and Mahevas (2007) for example). When the amount of data is large (as it will become with the implementation of the regional database) manual methods are (1) too time consuming and (2) are difficult to track in time (when and how the corrections were made). Hence automated procedures are needed to participate intrinsically to build the data quality.

This document is a tutorial related to the use of the R fishPifct package to assess data quality on fishery sampling data. The fishPifct package was developped for work package 4 of the fishPi project (project DG-MARE 2014/19 WP4). Its main objectives are to provide to the end users a framework to assess the quality of sampling data related to fishery.

3.4 Data format specification

This framework concerns sampling data and leans on the csPi format in term of data structure. The csPi format is a format under development based on the fishFrame format. The fishFrame format is used in the Regional Database and by the COST packages (a collection of tools to deal with data compilation, COST (2006)). Its complete definition is given in Jansen et al. (2009). An update of this format, called csPi, is in discussion since 2014 (ICES 2014), and the version 2.1 of this format is used in this report. Tools to export fishFrame files in csPi are provided, and in order to insure a wide range of application, most of the packages functions works on csPi and fishFrame objects.

3.5 Methodology

This framework follows the recommandation of the reproducible research statment (Gentleman and Lang 2004). Consequently this report is self-contained: the code used to process and analyse the sampling data are embedded in the report itself. An effort was made to select computing tools which give users the ability to reproduce the analyses using only a computer and an internet connection (for installation purposes mainly). Therefore all the tools are open source software, available free of charges, and running on the three main operating systems available (Linux, Windows and Mac OS).

3.6 Software

Coding and analyses are carried out using the R environment (R Core Team 2016). R[5] is a free software environment for statistical computing and graphics. The reproducibility of the results presented in this report relies on the use of a dialect of the Mardown language called Pandoc for word processing using the Knitr R package. Markdown is a plain text formatting syntax designed so that it can optionally be converted to HTML using a tool by the same name. Pandoc[6] is a Markdown dialect which extends the conversion capability to word processing file (docx, doc and odt), html and pdf, among other formats. Pandoc understands a number of markdown syntax extensions, including document metadata (title, author, date), footnotes, tables, figures and references. Knitr[7] is an R package (a set of functions extending the R capabilities). With this package, the R code used to process and analyze the data is included directly in the report. Results are then produced dynamically. This framework has demonstrated the capacity to improve the conduct and the presentation of data analysis in a way that another person can understand and replicate (Baumer et al. 2014).

For example, if the calculus of 1+1 is needed, the code to compute it is written in the report using special hooks, as in this simple example:

```
'''{r test00,warn=FALSE,cache=TRUE,echo=TRUE}
#comment: addition example.
1+1
'''
```

This code is evaluated during the compilation of the report by the knitr command and it prints the following result:

```
#comment: addition example.
1+1

## [1] 2
```

The result is 2. In this tutorial all the numerical values, tables and figures are produced following this procedure. The scripts and the report can be elaborated in a single integrated development environment (IDE), called Rstudio[8]. It includes a console and a syntax-highlighting editor that supports direct code execution, as well as tools for plotting, debugging and report writing. Consequently, all the tools and code presented here are already available to the end user.

[5] *http://www.r-project.org/*.
[6] *http://johnmacfarlane.net/pandoc/*.
[7] *http://yihui.name/knitr/*.
[8] *http://www.rstudio.com/*.

3.7 Installation

This package is available in the *fihPifct repository* on Github. The installation procedure is simple as:

```
install.packages("devtools")
library(devtools)
install_github("ldbk/fishPifct")
```

3.7.1 Issues

Technical support during the installation process (R version, missing packages, etc.) is far beyond the scope of this tutorial. In case of problems, please contact your IT support.

Some users reported issues with the openxlsx package installation (needed to import and export csPi and csData object in excel file). Please read carefully the error messages displayed by R (the way to fix these errors are generally explained in these messages). The standard procedure to fix errors should be something like:

```
install.packages("installr")
installr::installr("Rtools")
```

During the installation, tick the PATH modification option. Then, restart your computer.

3.8 COST library

If needed, COST related package (for windows) can be found here :

- *https://dl.dropboxusercontent.com/u/6181692/COSTcore_1.4-0.zip*
- *https://dl.dropboxusercontent.com/u/6181692/COSTdbe_1.4-1.zip*
- *https://dl.dropboxusercontent.com/u/6181692/COSTeda_1.4.0.zip*

and here for Unix system :

- *https://dl.dropboxusercontent.com/u/6181692/COSTcore_1.4-0.tar.gz*
- *https://dl.dropboxusercontent.com/u/6181692/COSTdbe_1.4-1.tar.gz*
- *https://dl.dropboxusercontent.com/u/6181692/COSTeda_1.4.0.tar.gz*

The COST user manual can be downloaded here :

- *https://dl.dropboxusercontent.com/u/6181692/COST%20User%20Manual%20V1_1.pdf*

4 Data

4.1 Format specification

In this tutorial, only the main characteristics of the csPi format are illustrated. A detailed version of the csPi format specifications is given in ICES (2014) and in the help page of the csPi function.

csPi is an S4 object containing 10 slots:

```
library(pander);library(fishPifct)
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
## Loading required package: openxlsx
## Loading required package: readxl
## Loading required package: rmarkdown
## Loading required package: robustbase
##
## Attaching package: 'robustbase'
## The following object is masked from 'package:stats':
##
##     sigma
## Loading required package: ggplot2
## Loading required package: mapdata
## Loading required package: maps
##
##  # maps v3.1: updated 'world': all lakes moved to separate new #
##  # 'lakes' database. Type '?world' or 'news(package="maps")'.  #
```

```
pander(format_definition_csPi$slots,split.table=Inf)
```

slot_name	mandatory	definition_table
classVersion	TRUE	slot_classVersion
desc	FALSE	slot_desc
popData	FALSE	slot_popData
design	FALSE	slot_design
se	TRUE	slot_se
tr	TRUE	slot_tr
hh	FALSE	slot_hh
sl	FALSE	slot_sl
hl	FALSE	slot_hl
ca	FALSE	slot_ca

The slots desc, popData, design are not mandatory and serve as descriptive fields for future applications.

The slots classVersion provides the version number of the csPi format. This format is still in development, and keeping the format version will insure retrocompatibility with the future development of the package. The slots hold the sampling information : the sampling events description (se), the trip information (tr), the hauls caracteristics (hh), the species sampled (sl), the correspondings length measurments (hl), and the biological parameters (ca). Each of these slots is a data.frame which lists the different parameters requested for each sample. The type of vessel, its characteristics, the fishing location and the quantity landed, the scientific name of the sampled species, the length class of the fishes, the age, etc. are reported in these tables. These variables can be numeric, text or codelist. For each table, a group of variables represent the primary key and insure the links with the other tables. The next figure gives an overview of the structure of the table.

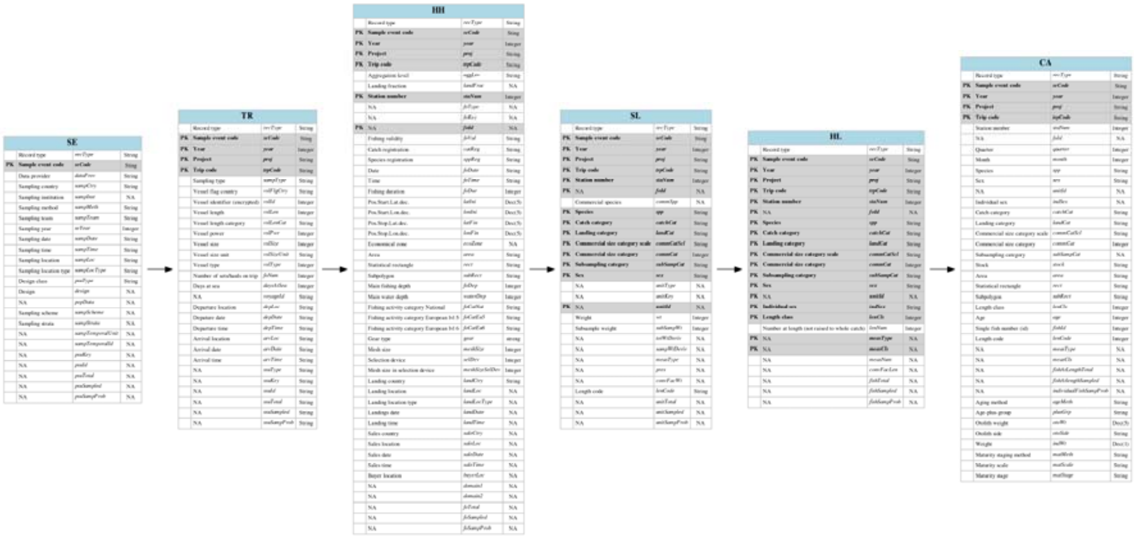


Figure 2. Overview of the csPi format

4.2 Example dataset

The data are generated based on the sole dataset coming from the COST package. The fishFrame COST format is exported in the csPi format using the function csDataTocsPi:

```
library(fishPifct)
data(sole)
sole <- csDataTocsPi(sole.cs)
## No seObj provided. Trips are used as sampling events. Use only for testing !
## Loading required package: COSTcore
## -------------------------------------------------------------------
##  COSTcore - "The DCF Statistical Suite"
##  (Version: 1.4-0. Built on: R 3.2.2; ; 2015-11-03 08:23:08 UTC; unix)
## -------------------------------------------------------------------
##
## Attaching package: 'COSTcore'
## The following object is masked from 'package:dplyr':
##
##     desc
head(sole)
## An object of class "csPi"
## Slot "classVersion":
## [1] "2.1"
##
## Slot "desc":
## [1] "Commercial Sampling Data format for the fishPi project"
##
## Slot "popData":
## [1] "Named population data object"
##
## Slot "design":
## [1] "Design description"
##
```

```
## Slot "history":
## [1] "modification history"
##
## Slot "se":
##   recType  seCode dataProv sampCtry sampInst sampMeth sampTeam seYear
## 1      se  ARY178      FRA      FRA   Obsmer       NA   Obsmer   2006
## 2      se DIL1196      FRA      FRA   Obsmer       NA   Obsmer   2006
## 3      se DIL1197      FRA      FRA   Obsmer       NA   Obsmer   2006
## 4      se  ELR214      FRA      FRA   Obsmer       NA   Obsmer   2006
## 5      se  ELR219      FRA      FRA   Obsmer       NA   Obsmer   2006
## 6      se   FAD73      FRA      FRA   Obsmer       NA   Obsmer   2006
##   sampDate sampTime sampLoc sampLocType psuType design popData sampScheme
## 1       NA       NA      NA          NA      NA     NA      NA         NA
## 2       NA       NA      NA          NA      NA     NA      NA         NA
## 3       NA       NA      NA          NA      NA     NA      NA         NA
## 4       NA       NA      NA          NA      NA     NA      NA         NA
## 5       NA       NA      NA          NA      NA     NA      NA         NA
## 6       NA       NA      NA          NA      NA     NA      NA         NA
##   sampStrata sampTemporalUnit sampTemporalId psuKey psuId psuTotal
## 1         NA               NA             NA     NA    NA       NA
## 2         NA               NA             NA     NA    NA       NA
## 3         NA               NA             NA     NA    NA       NA
## 4         NA               NA             NA     NA    NA       NA
## 5         NA               NA             NA     NA    NA       NA
## 6         NA               NA             NA     NA    NA       NA
##   psuSampled psuSampProb
## 1          1           1
## 2          1           1
## 3          1           1
## 4          1           1
## 5          1           1
## 6          1           1
##
## Slot "tr":
##   recType  seCode year   proj trpCode sampType vslFlgCtry vslId vslLen
## 1      tr  ARY178 2006 Obsmer  ARY178        S        FRA    98     NA
## 2      tr DIL1196 2006 Obsmer DIL1196        S        FRA    85     NA
## 3      tr DIL1197 2006 Obsmer DIL1197        S        FRA    21     NA
## 4      tr  ELR214 2006 Obsmer  ELR214        S        FRA    42     NA
## 5      tr  ELR219 2006 Obsmer  ELR219        S        FRA    43     NA
## 6      tr   FAD73 2006 Obsmer   FAD73        S        FRA    41     NA
##   vslLenCat vslPwr vslSize vslSizeUnit vslType foNum daysAtSea voyageId
## 1      <NA>     NA      NA        <NA>       1    27         4     <NA>
## 2      <NA>     NA      NA        <NA>       1    30         5     <NA>
## 3      <NA>     NA      NA        <NA>       1     5         2     <NA>
## 4      <NA>     NA      NA        <NA>       1    56        13     <NA>
## 5      <NA>     NA      NA        <NA>       1    13        10     <NA>
## 6      <NA>     NA      NA        <NA>       3     3         4     <NA>
##   depLoc depDate depTime arvLoc arvDate arvTime ssuType ssuKey ssuId
## 1   <NA>    <NA>    <NA>   <NA>    <NA>    <NA>    <NA>   <NA>  <NA>
## 2   <NA>    <NA>    <NA>   <NA>    <NA>    <NA>    <NA>   <NA>  <NA>
## 3   <NA>    <NA>    <NA>   <NA>    <NA>    <NA>    <NA>   <NA>  <NA>
```

```
## 4  <NA>   <NA>   <NA>  <NA>   <NA>   <NA>   <NA>  <NA> <NA>
## 5  <NA>   <NA>   <NA>  <NA>   <NA>   <NA>   <NA>  <NA> <NA>
## 6  <NA>   <NA>   <NA>  <NA>   <NA>   <NA>   <NA>  <NA> <NA>
##   ssuTotal ssuSampled ssuSampProb
## 1     <NA>       <NA>        <NA>
## 2     <NA>       <NA>        <NA>
## 3     <NA>       <NA>        <NA>
## 4     <NA>       <NA>        <NA>
## 5     <NA>       <NA>        <NA>
## 6     <NA>       <NA>        <NA>
##
## Slot "hh":
##   recType seCode year   proj trpCode aggLev landFrac staNum foType foKey
## 1      hh ARY178 2006 Obsmer  ARY178      H     <NA>      1   <NA>  <NA>
## 2      hh ARY178 2006 Obsmer  ARY178      H     <NA>      2   <NA>  <NA>
## 3      hh ARY178 2006 Obsmer  ARY178      H     <NA>      3   <NA>  <NA>
## 4      hh ARY178 2006 Obsmer  ARY178      H     <NA>      4   <NA>  <NA>
## 5      hh ARY178 2006 Obsmer  ARY178      H     <NA>      5   <NA>  <NA>
## 6      hh ARY178 2006 Obsmer  ARY178      H     <NA>      6   <NA>  <NA>
##   foId foVal catReg sppReg     foDate foTime foDur   latIni   lonIni
## 1 <NA>     V    All    All 2006-04-03   <NA>   150 50.05360 1.623667
## 2 <NA>     V    All    All 2006-04-03   <NA>   180 51.19133 1.786333
## 3 <NA>     V    Non    Non 2006-04-03   <NA>   165 51.11667 1.672667
## 4 <NA>     V    All    All 2006-04-03   <NA>   195 51.03933 1.568667
## 5 <NA>     V    All    All 2006-04-03   <NA>   190 51.03933 1.666667
## 6 <NA>     V    Non    Non 2006-04-04   <NA>   180 51.11667 1.672667
##   latFin lonFin ecoZone   area rect subRect foDep waterDep     foCatNat
## 1     NA     NA    <NA> 27.7.d 29F1    <NA>    NA       40 7D-OTB-Merlan
## 2     NA     NA    <NA> 27.4.c 31F1    <NA>    NA       40 4C-OTB-Merlan
## 3     NA     NA    <NA> 27.4.c 31F1    <NA>    NA       40 4C-OTB-Merlan
## 4     NA     NA    <NA> 27.4.c 31F1    <NA>    NA       40 4C-OTB-Merlan
## 5     NA     NA    <NA> 27.4.c 31F1    <NA>    NA       40 4C-OTB-Merlan
## 6     NA     NA    <NA> 27.4.c 31F1    <NA>    NA       40 4C-OTB-Merlan
##   foCatEu5       foCatEu6 gear meshSize selDev meshSizeSelDev landCtry
## 1  OTB_DEF OTB_DEF_80_0_0 <NA>       80      0             NA      FRA
## 2  OTB_DEF OTB_DEF_80_0_0 <NA>       80      0             NA      FRA
## 3  OTB_DEF OTB_DEF_80_0_0 <NA>       80      0             NA      FRA
## 4  OTB_DEF OTB_DEF_80_0_0 <NA>       80      0             NA      FRA
## 5  OTB_DEF OTB_DEF_80_0_0 <NA>       80      0             NA      FRA
## 6  OTB_DEF OTB_DEF_80_0_0 <NA>       80      0             NA      FRA
##   landLoc landLocType landDate landTime saleCtry saleLoc saleDate saleTime
## 1    <NA>        <NA>     <NA>     <NA>     <NA>    <NA>     <NA>     <NA>
## 2    <NA>        <NA>     <NA>     <NA>     <NA>    <NA>     <NA>     <NA>
## 3    <NA>        <NA>     <NA>     <NA>     <NA>    <NA>     <NA>     <NA>
## 4    <NA>        <NA>     <NA>     <NA>     <NA>    <NA>     <NA>     <NA>
## 5    <NA>        <NA>     <NA>     <NA>     <NA>    <NA>     <NA>     <NA>
## 6    <NA>        <NA>     <NA>     <NA>     <NA>    <NA>     <NA>     <NA>
##   buyerLoc domain1 domain2 foTotal foSampled foSampProb
## 1     <NA>    <NA>    <NA>    <NA>      <NA>       <NA>
## 2     <NA>    <NA>    <NA>    <NA>      <NA>       <NA>
## 3     <NA>    <NA>    <NA>    <NA>      <NA>       <NA>
## 4     <NA>    <NA>    <NA>    <NA>      <NA>       <NA>
```

```
## 5    <NA>    <NA>    <NA>    <NA>     <NA>      <NA>
## 6    <NA>    <NA>    <NA>    <NA>     <NA>      <NA>
##
## Slot "sl":
##   recType  seCode year   proj trpCode staNum foId commSpp        spp
## 1      sl DIL1197 2006 Obsmer DIL1197      1 <NA>    <NA> Solea solea
## 2      sl DIL1197 2006 Obsmer DIL1197      1 <NA>    <NA> Solea solea
## 3      sl DIL1197 2006 Obsmer DIL1197      2 <NA>    <NA> Solea solea
## 4      sl DIL1197 2006 Obsmer DIL1197      2 <NA>    <NA> Solea solea
## 5      sl DIL1197 2006 Obsmer DIL1197      3 <NA>    <NA> Solea solea
## 6      sl DIL1197 2006 Obsmer DIL1197      3 <NA>    <NA> Solea solea
##   catchCat landCat commCatScl commCat subSampCat  sex unitType unitKey
## 1      LAN     HUC         EU    <NA>       <NA> <NA>     <NA>    <NA>
## 2      DIS     HUC         EU    <NA>       <NA> <NA>     <NA>    <NA>
## 3      LAN     HUC         EU    <NA>       <NA> <NA>     <NA>    <NA>
## 4      DIS     HUC         EU    <NA>       <NA> <NA>     <NA>    <NA>
## 5      LAN     HUC         EU    <NA>       <NA> <NA>     <NA>    <NA>
## 6      DIS     HUC         EU    <NA>       <NA> <NA>     <NA>    <NA>
##   unitId    wt subSampWt totWtDeriv sampWtDeriv measType pres convFacWt
## 1   <NA> 11000        NA       <NA>        <NA>     <NA> <NA>      <NA>
## 2   <NA> 10000      1560       <NA>        <NA>     <NA> <NA>      <NA>
## 3   <NA> 26321        NA       <NA>        <NA>     <NA> <NA>      <NA>
## 4   <NA> 12000      2570       <NA>        <NA>     <NA> <NA>      <NA>
## 5   <NA> 73000        NA       <NA>        <NA>     <NA> <NA>      <NA>
## 6   <NA>  7000      4217       <NA>        <NA>     <NA> <NA>      <NA>
##   lenCode unitTotal unitSampled unitSampProb
## 1      cm      <NA>        <NA>         <NA>
## 2      cm      <NA>        <NA>         <NA>
## 3      cm      <NA>        <NA>         <NA>
## 4      cm      <NA>        <NA>         <NA>
## 5      cm      <NA>        <NA>         <NA>
## 6      cm      <NA>        <NA>         <NA>
##
## Slot "hl":
##   recType  seCode year   proj trpCode staNum foId         spp catchCat
## 1      hl DIL1197 2006 Obsmer DIL1197      1 <NA> Solea solea      DIS
## 2      hl DIL1197 2006 Obsmer DIL1197      1 <NA> Solea solea      DIS
## 3      hl DIL1197 2006 Obsmer DIL1197      1 <NA> Solea solea      DIS
## 4      hl DIL1197 2006 Obsmer DIL1197      1 <NA> Solea solea      DIS
## 5      hl DIL1197 2006 Obsmer DIL1197      2 <NA> Solea solea      DIS
## 6      hl DIL1197 2006 Obsmer DIL1197      2 <NA> Solea solea      DIS
##   landCat commCatScl commCat subSampCat  sex unitId indSex lenCls lenNum
## 1     HUC         EU    <NA>       <NA> <NA>   <NA>   <NA>    180      3
## 2     HUC         EU    <NA>       <NA> <NA>   <NA>   <NA>    190      6
## 3     HUC         EU    <NA>       <NA> <NA>   <NA>   <NA>    200      5
## 4     HUC         EU    <NA>       <NA> <NA>   <NA>   <NA>    210      1
## 5     HUC         EU    <NA>       <NA> <NA>   <NA>   <NA>    160      1
## 6     HUC         EU    <NA>       <NA> <NA>   <NA>   <NA>    170      1
##   measType measCls measNum convFacLen fishTotal fishSampled fishSampProb
## 1     <NA>    <NA>    <NA>       <NA>      <NA>        <NA>         <NA>
## 2     <NA>    <NA>    <NA>       <NA>      <NA>        <NA>         <NA>
## 3     <NA>    <NA>    <NA>       <NA>      <NA>        <NA>         <NA>
```

```
## 4    <NA>   <NA>   <NA>     <NA>     <NA>      <NA>       <NA>
## 5    <NA>   <NA>   <NA>     <NA>     <NA>      <NA>       <NA>
## 6    <NA>   <NA>   <NA>     <NA>     <NA>      <NA>       <NA>
##
## Slot "ca":
##   recType seCode year   proj trpCode staNum foId quarter month        spp
## 1      ca     12 2006 BioPar      12    999 <NA>       2     4 Solea solea
## 2      ca     12 2006 BioPar      12    999 <NA>       2     4 Solea solea
## 3      ca     12 2006 BioPar      12    999 <NA>       2     4 Solea solea
## 4      ca     12 2006 BioPar      12    999 <NA>       2     4 Solea solea
## 5      ca     12 2006 BioPar      12    999 <NA>       2     4 Solea solea
## 6      ca     12 2006 BioPar      12    999 <NA>       2     4 Solea solea
##   sex unitId indSex catchCat landCat commCatScl commCat subSampCat stock
## 1   M   <NA>   <NA>      LAN     HUC       <NA>    <NA>       <NA>  <NA>
## 2   M   <NA>   <NA>      LAN     HUC       <NA>    <NA>       <NA>  <NA>
## 3   M   <NA>   <NA>      LAN     HUC       <NA>    <NA>       <NA>  <NA>
## 4   F   <NA>   <NA>      LAN     HUC       <NA>    <NA>       <NA>  <NA>
## 5   F   <NA>   <NA>      LAN     HUC       <NA>    <NA>       <NA>  <NA>
## 6   M   <NA>   <NA>      LAN     HUC       <NA>    <NA>       <NA>  <NA>
##     area rect subRect lenCls age fishId lenCode measType measCls
## 1 27.7.d 28E9    <NA>    330   5      1      cm     <NA>    <NA>
## 2 27.7.d 28E9    <NA>    340   7      2      cm     <NA>    <NA>
## 3 27.7.d 28E9    <NA>    320   7      3      cm     <NA>    <NA>
## 4 27.7.d 28E9    <NA>    320   4      4      cm     <NA>    <NA>
## 5 27.7.d 28E9    <NA>    350   7      5      cm     <NA>    <NA>
## 6 27.7.d 28E9    <NA>    340   9      6      cm     <NA>    <NA>
##   fishAtLengthTotal fishAtlengthSampled individualFishSampProb
## 1              <NA>                <NA>                   <NA>
## 2              <NA>                <NA>                   <NA>
## 3              <NA>                <NA>                   <NA>
## 4              <NA>                <NA>                   <NA>
## 5              <NA>                <NA>                   <NA>
## 6              <NA>                <NA>                   <NA>
##                              ageMeth plusGrp otoWt otoSide indWt
## 1 Otoliths - slides with transmitted light       -    NA    <NA>   355
## 2 Otoliths - slides with transmitted light       -    NA    <NA>   360
## 3 Otoliths - slides with transmitted light       -    NA    <NA>   339
## 4 Otoliths - slides with transmitted light       -    NA    <NA>   416
## 5 Otoliths - slides with transmitted light       -    NA    <NA>   412
## 6 Otoliths - slides with transmitted light       -    NA    <NA>   411
##   matMeth matScale matStage
## 1  Visual      1-7        2
## 2  Visual      1-7        2
## 3  Visual      1-7        2
## 4  Visual      1-7        5
## 5  Visual      1-7        4
## 6  Visual      1-7        2
```

The csPi object is named sole in our example.

4.3 Handling csPi objects

A collection of methods and functions gives to the user the ability to explore and visualize a csPi objects:

```
methods(class="csPi")
## [1] dim     export  head    summary tail
## see '?methods' for accessing help and source code
#subset csPi object
#?csSubset
```

Their behaviours are similar to the generic one (i.e. dim gives the dimension of all the csPi slots).

4.4 Import and export in spreadsheet

Fishing data relies usually on national databases. Correction procedures in these systems can be a tedious work, not really in accordance to quick corrections (during working groups, to harmonize datasets between countries for example). Manual data corrections are difficult and spreadsheet is nowadays the common tools to correct locally the data. A local import/export procedure is available to export the csPi in excel file format. Thus, the user can use a spreadsheet to do some corrections in the tables and then import directly the corrected tables in a csPi object in R.

In this package the import and export functions do these transformation easily:

```
export(sole,file="sole.xlsx",type="xlsx")
## [1] "sole.xlsx"
#use a spreadsheet to open the sole.xlsx file and do some corrections if needed
#save the file, and import it in R with:
solecorrected<-importxlsx(file="sole.xlsx")
```

5 Data quality checks

5.1 Data structure checks

A seminal step in data quality is to check the structure of the data. The structure check includes the ordered verification of :

- the objects' slots: name, existence, mandatory or not.
- the slots' tables: dimension, variables names, mandatory or not, uniqueness of the primary keys if applicable.
- the tables' variables: their types - numeric (integer or real, lower and upper limits), character, codelist (a list of authorized values)-, nullable, mandatory or not.

The data structure definition is given for csPi objects by the list format_definition_csPi. This list is built from the excel file format_definition_csPi.xlsx available in the data directory of the installation directory of the package and detailed in Annex A. The excel file is the unique support to the entire data check, which means that RCG may agree on all the references for a regional database solely by defining every code lists and numeric boundaries (for example the lower and upper limits of the length class, or a limited list of métier) of the format definition excel file, and these will become the references upon which any dataset will be checked.

5.1.1 Slot definition

A slot definition is a table reporting the characteristics of a given slot :

slot_name	mandatory	definition_table
base	TRUE	slot_base

Here the slot names base is mandatory and its definition is given by the table slot_base. During the structure check, each slot is checked against its definition given by the structure definition list.

5.1.2 Table definition

A table definition is a table reporting the characteristics of a given table. For example here, the first 8 lines of the tr table definition :

```
library(pander);library(fishPifct)
pander(format_definition_csPi$slot_tr[1:8,],split.table=Inf)
```

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	FALSE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
year	FALSE	TRUE	TRUE	type_year	numeric
proj	FALSE	TRUE	TRUE	type_proj	text
trpCode	FALSE	TRUE	TRUE	type_trpCode	text
sampType	FALSE	TRUE	FALSE	type_sampType	codelist
vslFlgCtry	FALSE	TRUE	FALSE	type_ctry	codelist
vslId	FALSE	FALSE	FALSE	type_vslId	text

Each table's column is checked against its definition. For example, the trpCode variable has to be non nullable, is mandatory and is part of the primary key of the tr table. It is a text variable (category), and its category definition is referenced in the type_trpCode of the definition file (or the excel sheet with this name).

5.1.3 Variable checks

After the table definition, each variable is checked according to their types. For example in the previous table vslFlagCtry is non nullable, mandatory and is not included in the primary key. The variable's type is a codelist, and the corresponding authorized value is registered in the codelist_type list of the format description, namely the list codelist_ctry (here the first 10 lines):

```
pander(format_definition_csPi$codelist_ctry[1:10,],split.table=Inf)
```

CODE	DESCRIPTION
ABW	Aruba
AFG	Afghanistan
AGO	Angola
AIA	Anguilla
ALA	Åland Islands
ALB	Albania
AND	Andorra
ARE	United Arab Emirates
ARG	Argentina
ARM	Armenia

This list is the list of the ISO 3166-1 alpha-3 country codes. Limiting this list strengthens the data quality check, according to the end user needs.

For those variables of a numeric type, the numeric_type list of the data definition brings information related to the numerical limits and if the numbers are integer (number of samples, age, etc.) or real (probability).
For example here, the first 8 lines of the numeric_type table definition:

```
library(pander);library(fishPifct)
pander(format_definition_csPi$numeric_type[1:8,],split.table=Inf)
```

type_name	is_integer	min	max
type_year	TRUE	1900	2020
type_psuTotal	TRUE	0	1e+07
type_psuSampled	TRUE	0	2000
type_psuSampProb	FALSE	0	1
type_vslLen	TRUE	3	160
type_vslPwr	TRUE	4	8500
type_vslSize	TRUE	1	2500
type_foNum	TRUE	1	300

In this example, year is an integer between 1900 and 2020. As previously stated the modification of the data structure is open to the end user needs.

5.1.4 Notes

The data structure checks were developped by the sister project of fishPi related to the Mediterranean area, for fishFrame object (*https://git.outils-is.ird.fr/billet/SDEFQuality/wikis/home*). Consequently, this data structure check is applicable to any object structure, and it can be extended to landings or effort file in a near future for example.

5.1.5 Outputs

The results of the data structure checks are given in a report summarising all the checks, if these checks pass, and why. Using the sole dataset previously loaded:

```
#generating a report in an R object
structurecheck<-validateData(obj=sole,formatDb=format_definition_csPi,report="list")
```

The meta information related to the check are:

```
pander(structurecheck$meta,split.table=Inf)
```

parameter	value
format_name	csPi
format_version	2.1
validate_date	2016-05-19 09:04:26
dataset_container	object
format_container	object

The 10 first lines of the slots checks are:

```
pander(structurecheck$struct[1:10,],split.table=Inf)
```

slot	column	test	result	message

classVersion	NA	Slot exists ?	OK	Found
classVersion	classVersion	Column exists ?	OK	Found
desc	NA	Slot exists ?	OK	Found
desc	desc	Column exists ?	OK	Found
popData	NA	Slot exists ?	OK	Found
popData	popData	Column exists ?	OK	Found
design	NA	Slot exists ?	OK	Found
design	design	Column exists ?	OK	Found
se	NA	Slot exists ?	OK	Found
se	recType	Column exists ?	OK	Found

The 10 first lines of the variables checks are:

```
pander(structurecheck$data[1:10,],split.table=Inf)
```

slot	column	test	result	message
classVersion	classVersion	is text ?	OK	All values are text
classVersion	classVersion	is null ?	OK	All values are not null
desc	desc	is text ?	OK	All values are text
desc	desc	is null ?	OK	All values are not null
popData	popData	is text ?	OK	All values are text
popData	popData	is null ?	OK	All values are not null
design	design	is text ?	OK	All values are text
design	design	is null ?	OK	All values are not null
se	recType	is valid code list ?	OK	All values are valid codes
se	recType	is null ?	OK	All values are not null

The tables are explicits and do not need any comments. To generate a complete report in pdf or html format :

```
#generating a pdf report
renderValidationReport(obj=sole,formatDb=format_definition_csPi,
        title="test",reportFormat="pdf")
```

Report generated [/tmp/RtmpHXMPUo/dataValidationReport_20160519_090427_17163132ffce.pdf]

[1] "/tmp/RtmpHXMPUo/dataValidationReport_20160519_090427_17163132ffce.pdf"

```
#a copy of this report can be found in
system.file('data',
'dataValidationReport_20160518_235140_29d51c808f9b.pdf',
     package='fishPifct')
```

[1] "/home/moi/R/x86_64-pc-linux-gnu-library/3.3/fishPifct/data/dataValidationReport_20160518_235140_29d51c808f9b.pdf"

5.1.6 Consistency check

In this section, the consistency of the information between the 'csPi' slots is checked, e.g. identification of trips without fishing operations. To do so, the function consistency performs hierachical anti jointure between related table and generates a simple table reporting the trpCode which has to be checked between the tables (see Annex B)

```
#consistency check generating a pdf report
consistencycheck<-consistency(sole)
pander(consistencycheck,split.table=Inf)
```

test	message	check
tr->se	0 tr records have no correspondings se records	orphans tr trpCode:
hh->tr	0 hh records have no correspondings tr records	orphans hh trpCode:
sl->hh	0 sl records have no correspondings hh records	orphans sl trpCode:
hl->sl	0 hl records have no correspondings sl records	orphans hl trpCode:

In our sole example, no consistency errors were detected.

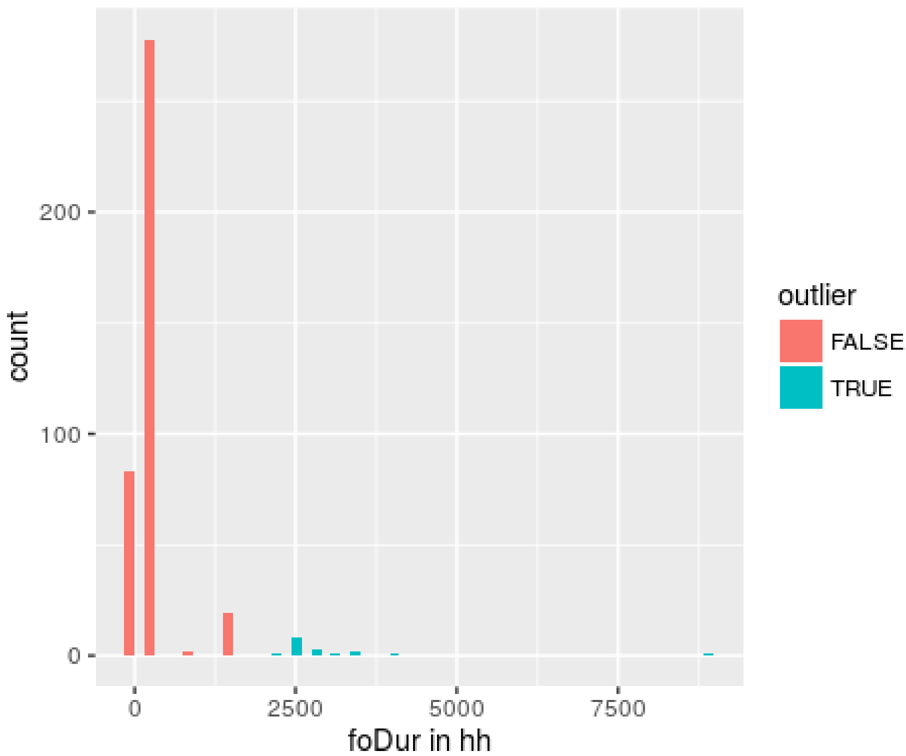
5.2 Outliers detection

The literature on outliers is extensive, and cover all the areas of science, but determining whether or not an observation is an outlier is ultimately a subjective exercise and hence makes automation a difficult task. Here we will use the definition of Barnett and Lewis (1994) for outlier: "Indicate that an outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs". Outlier detection methods can be divide between univariate methods (looking at only one variable) and multivariate methods (looking at more than one variable and their relationships). For example univariate methods spot observations reported in tons instead of kilos in landings, while multivariate methods can identify wrong weigths in a size-weight relationship. Then outlier detection methods can be categorized between parametric (statistical) methods and non-parametric methods that are model free. Statistical parametric methods either assume a known underlying distribution of the data or, at least, they are based on statistical estimates of unknown distribution parameters. Observations that deviate from the model assumptions are flagged as outliers. Here we focus on two generic non parametric methods for numerical and non numerical univariate data. The function outliers do the outliers detection for a csPi object.

5.2.1 Numeric variables

The adjusted outlyingness index is used to detect outliers. It's a non parametric method, adapted to skewed data. The function adjOutliness of the package robustBase is used. More details of this method can be found in the help page of this function. An example on the fishing duration (variable foDur of slot hh):

```
tabaoutlier<-outliers(sole,slot="hh",var="foDur")
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 48 rows containing non-finite values (stat_bin).
```



High fishing duration values are flagged as outliers, as presented in the figure. The function output gives to the user the complete lines who includes the outliers:

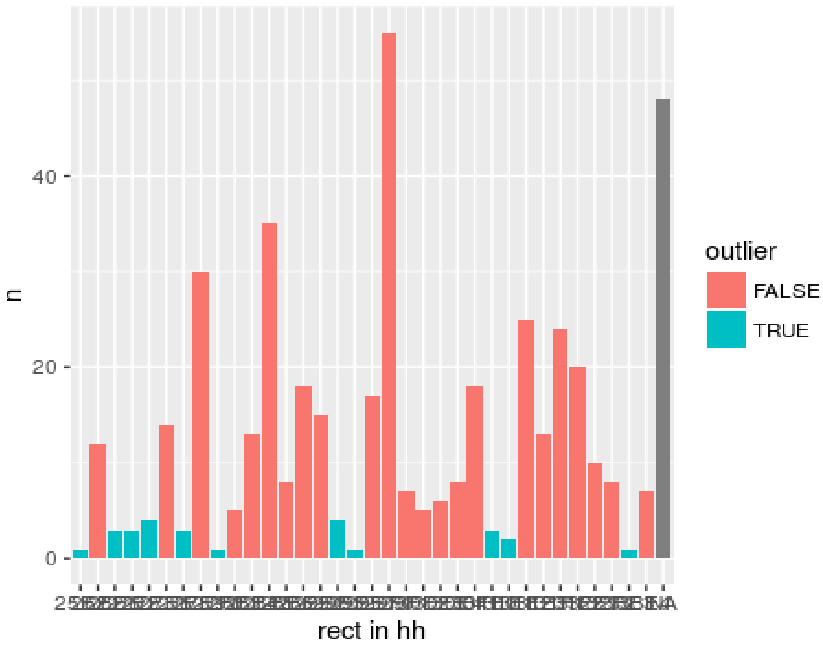
```
#10 first lines and 5 first columns of the outliers
pander(tabaoutlier[1:10,1:5],split.table=Inf)
```

	recType	seCode	year	proj	trpCode
63	hh	ELR214	2006	Obsmer	ELR214
274	hh	MAC3	2006	Obsmer	MAC3
275	hh	MAC3	2006	Obsmer	MAC3
276	hh	MAC3	2006	Obsmer	MAC3
277	hh	MAC3	2006	Obsmer	MAC3
278	hh	MAC3	2006	Obsmer	MAC3
279	hh	MAC3	2006	Obsmer	MAC3
280	hh	MAC3	2006	Obsmer	MAC3
281	hh	MAC3	2006	Obsmer	MAC3
282	hh	MAC3	2006	Obsmer	MAC3

5.2.2 Text and codelist variables

For a non-numerical variable, the outliers are detected using the occurence of the modality of the value, expressed in percentage and a threshold (by default 1%). If a modality is expressed less than this threshold, an outlier is considered detected. The threshold can be fixed by the user. The figure below provides an example using the statistical rectangle fished :

```
tabaoutlier<-outliers(sole,slot="hh",var="rect")
```



Rare fished rectangle are flagged as outliers. The function output gives to the user the complete lines who includes the outliers:

```
#10 first lines and 5 first columns of the outliers
pander(tabaoutlier[1:10,1:5],split.table=Inf)
```

	recType	seCode	year	proj	trpCode
65	hh	ELR214	2006	Obsmer	ELR214
68	hh	ELR214	2006	Obsmer	ELR214
77	hh	ELR214	2006	Obsmer	ELR214
78	hh	ELR214	2006	Obsmer	ELR214
79	hh	ELR214	2006	Obsmer	ELR214
81	hh	ELR214	2006	Obsmer	ELR214
82	hh	ELR214	2006	Obsmer	ELR214
83	hh	ELR214	2006	Obsmer	ELR214
90	hh	ELR214	2006	Obsmer	ELR214
110	hh	ELR214	2006	Obsmer	ELR214

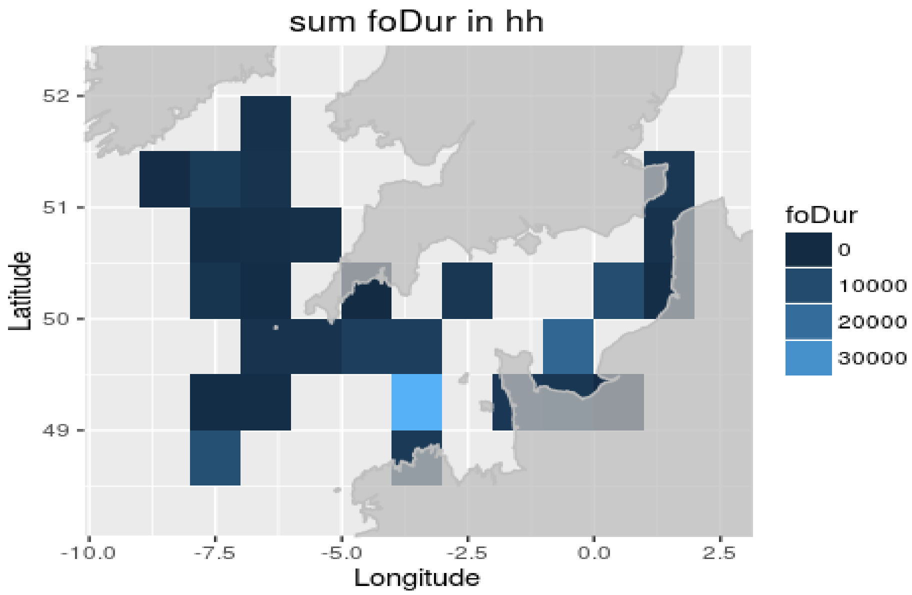
These functions are generic and can be applied to the whole set of variables of a csPi object.

5.3 Plots

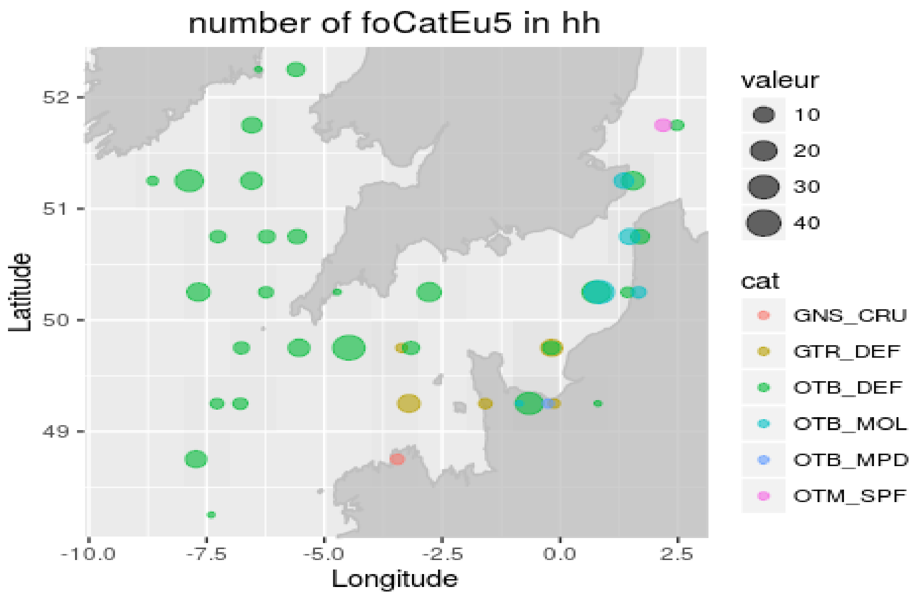
5.3.1 Maps

The function csMap gives the ability to the user to map easily a variable over the ICES statistical rectangle.

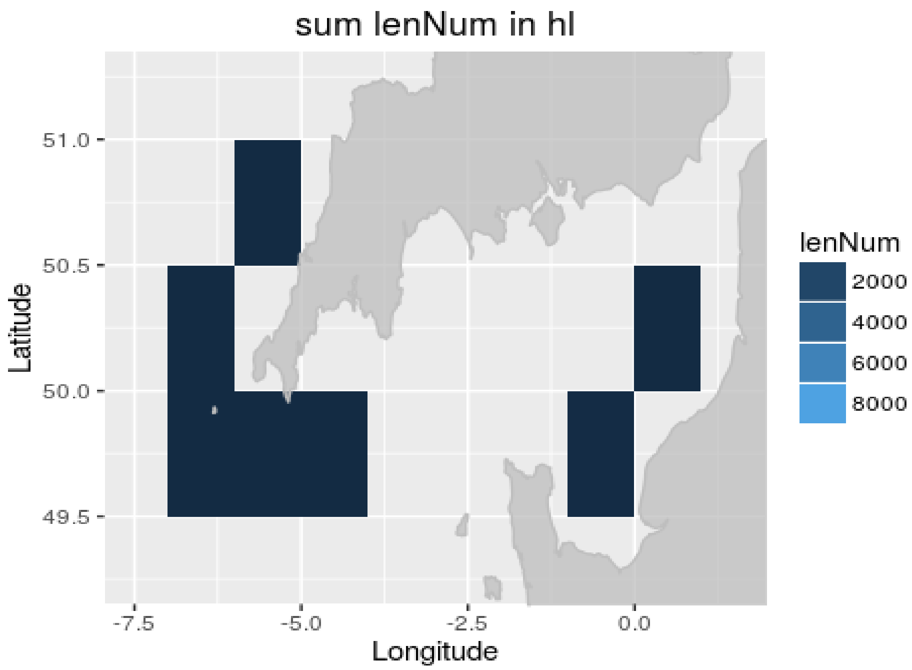
```
csMap(sole,"hh","foDur","tile","sum")
## [1] "Statistical rectangles missing and not mapped: "
## Joining by: "rect"
## Warning: Removed 1 rows containing missing values (geom_tile).
```



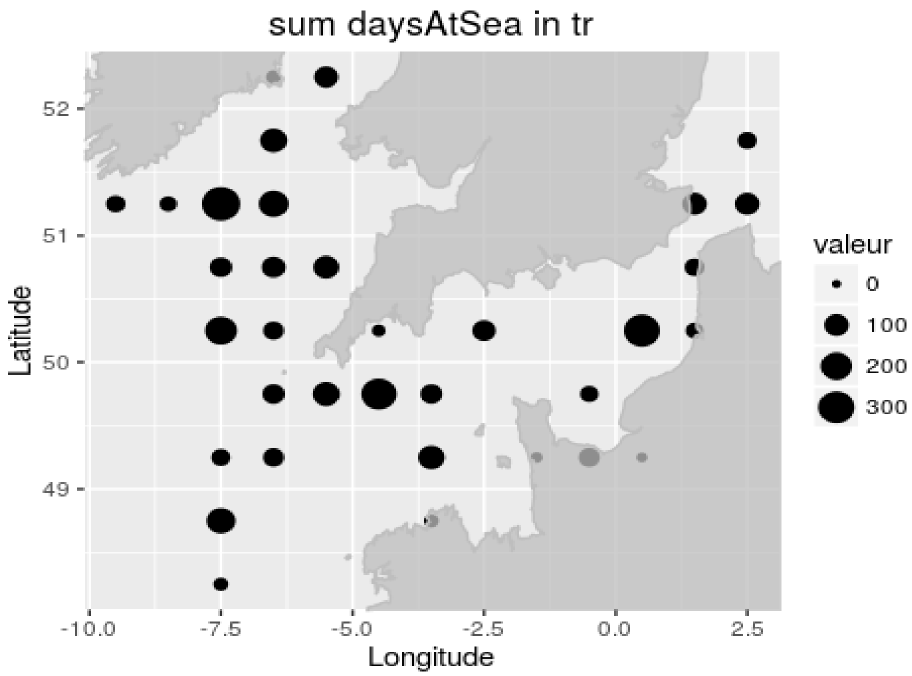

```
csMap(sole,"hh","foCatEu5","bubble")
## [1] "Statistical rectangles missing and not mapped: "
## Joining by: "rect"
## Warning: Removed 1 rows containing missing values (geom_tile).
## Warning: Removed 3 rows containing missing values (geom_point).
```



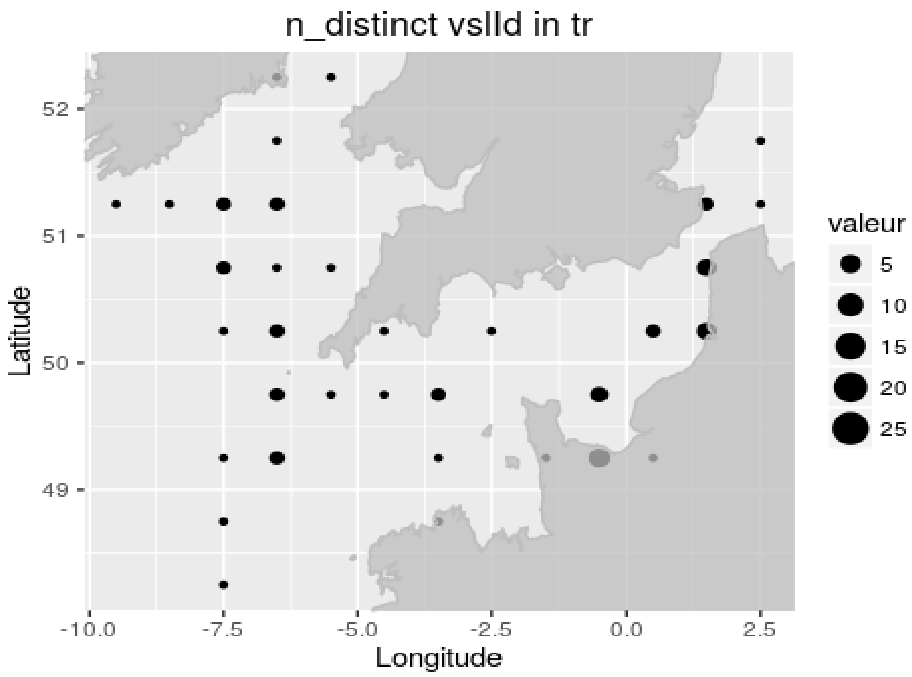

```
csMap(sole,"hl","lenNum","tile","sum")
## [1] "Statistical rectangles missing and not mapped: "
## Joining by: "rect"
## Warning: Removed 1 rows containing missing values (geom_tile).
```




```
csMap(sole,"tr","daysAtSea","bubble","sum")
## [1] "Statistical rectangles missing and not mapped: "
## Joining by: "rect"
## Warning: Removed 1 rows containing missing values (geom_point).
```



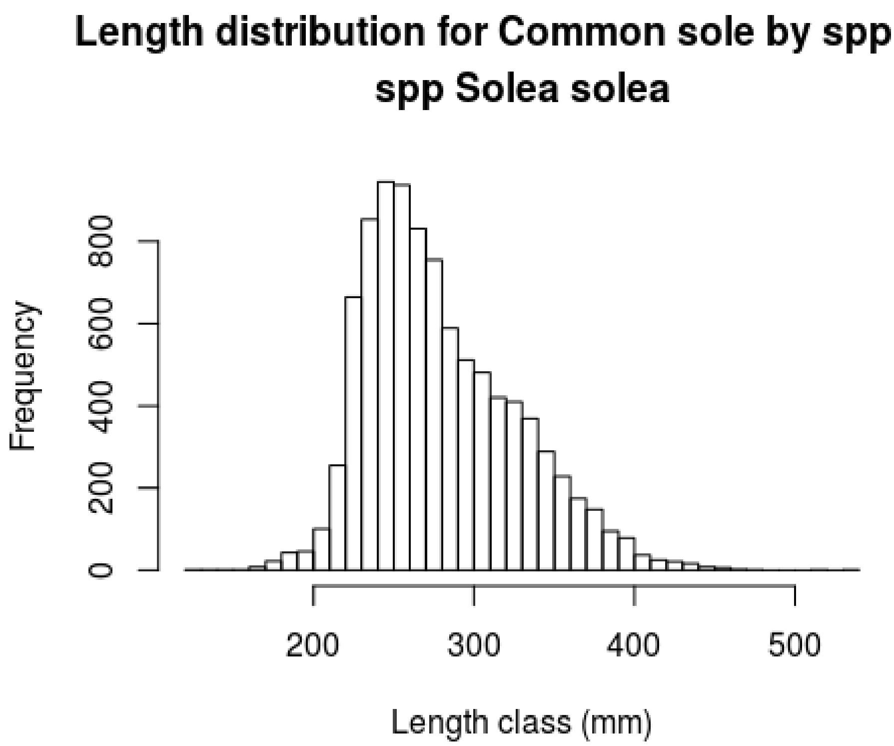

```
csMap(sole,"tr","vslId","bubble","n_distinct")

## [1] "Statistical rectangles missing and not mapped: "
## Joining by: "rect"
## Warning: Removed 1 rows containing missing values (geom_point).
```



5.3.2 Generic functions

The function lengthHist plots histograms of the length frequency data from the hl table of csPi object.

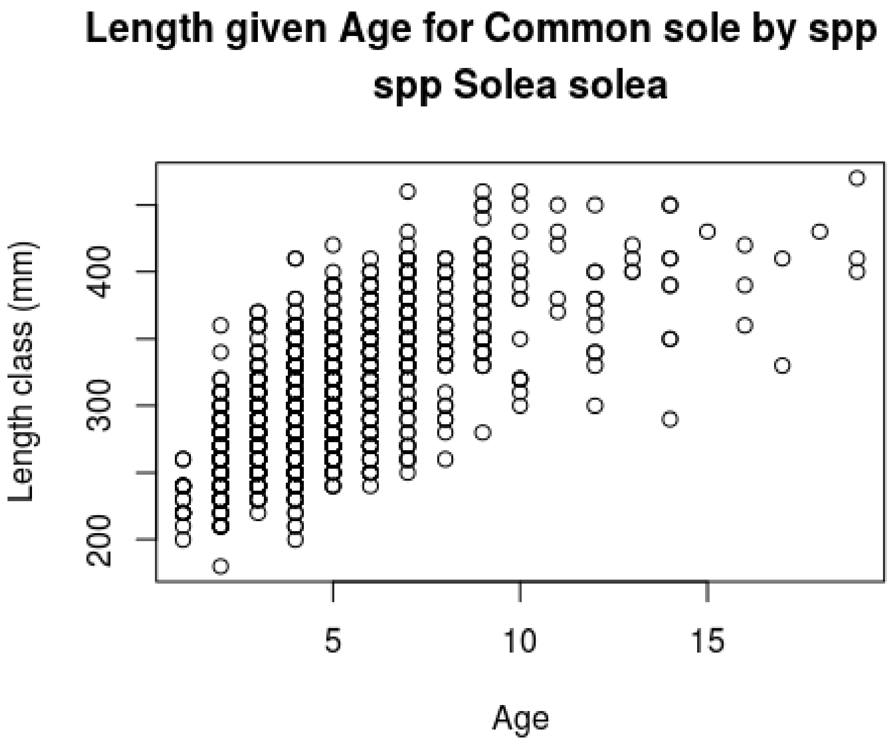
```
lengthHist(sole)
```



The function agelenPlot plots age given length from the ca table of a csPi object.

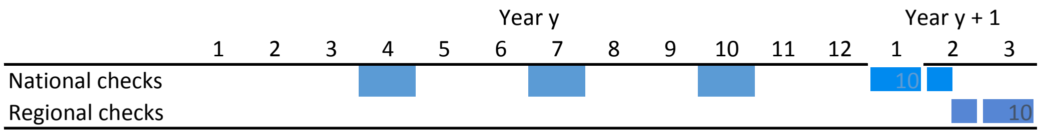
```
agelenPlot(sole)
```

Warning in agelenPlot(sole): Only LAN fraction present in data



5.3.3 Calendar for applying the QC at a national and regional level

In 2016, the earliest dates for submitting data to ICES were determined by WGFBAS and HAWG assessment working groups respectively on 15 and 16 March 2016. In order to allocate time to validate and process the data regionally and meet these deadlines, the national data should be validated and checked nationally and uploaded to the RDB before 15 February. To ease this process, Member States should run the quality checks and validate national data on a quarterly basis, as shown on the schema below.



However, this has to be considered as an objective calendar. Until automatic procedures, like those promoted in this work package, are implemented routinely in each country, it is deemed impossible to respect these deadlines. Moreover, some countries expressed the need to be more flexible in the national quarterly checks, in order to accomodate national constraints. Eventually, a step by step procedure is demanded taking into account the different deadlines of the ICES datacalls.

6 Discussion

Key developments were made on data handling the new data structure and quality checks during the project, beyond the coding of the R functions. While the proposed R functions provide the mechanisms and process for data handling and quality checks the principal behind these can be summarised as follows:

- The RDB data format is not settled yet, and may be modified without affecting the data handling and quality check functions;
- The data structure and all references (table names, field names and types, field boundaries, codelists) are defined in an Excel file (Annex A), which can be agreed and maintained by the RCG. This (regional) reference excel file should be made available on a public website, in order to be used nationally for any international data exchange;
- The R functions are using the Excel file as support to quality check the tables prior to import. The R functions are available on a public website, and may be improved on a continuous basis;
- The output of the R functions is also an excel file (Annex B), containing each of the imported tables which are augmented with an extra column. The extra column contains any errors and description of issues positioned at the very level of the suspicious record. This method is made for easing validation and correction work, and can be used either at national or regional level.

All the elements detailed above allow for quality checks to be run at national level on a routine basis, to ensure that the most qualified data is inputted into national databases. Given that the references used to test the datasets are the same at national and regional levels, the qualified national data may be uploaded into regional databases, where quality checks can qualify the regional datasets.

In order to implement gradually the proposed approach, a roadmap for a short term development plan and the resources needed are detailed in the work programme section of WP1.

6.1 About this vignette

This vignette was built using the vignette engine knitr::rmarkdown in the **knitr** package. The source can be found in the *fihPifct repository* on Github, or if the fishPifct package is installed on your computer:

```
system.file('doc', 'tutorial.Rmd', package='fishPifct')
## [1] "/home/moi/R/x86_64-pc-linux-gnu-library/3.3/fishPifct/doc/tutorial.Rmd"
```

References

Barnett, V, and T Lewis. 1994. *Outliers in Statistical Data. 3rd Ed.* Wiley.

Baumer, Ben, Mine Cetinkaya-Rundel, Andrew Bray, Linda Loi, and Nicholas J. Horton. 2014. "R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics." *Technology Innovations in Statistics Education* 8 (1).

Ben-Gal I. 2005. Outlier detection, In: Maimon O. and Rockach L. (Eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Kluwer Academic Publishers, 2005.

Chen, Y. 2003. "Quality of Fisheries Data and Uncertainty in Stock Assessment." *Scientia Marina* 67 (1): 75–87.

COST. 2006. *Studies and Pilot Projects for Carrying Out the Common Fisheries Policy Call for Proposal Ref : FISH/2006/15 – Lot 2 Project No :SI2.467814 Project Acronym :COST*. EUROPEAN COMMISSION.

Gentleman, Robert, and Duncan Temple Lang. 2004. "Statistical Analyses and Reproducible Research." Bioconductor Project Working Papers. Working Paper 2. *http://www.bepress.com/bioconductor/paper2*.

ICES 2008. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU). ICES CM 2008/ACOM:32, 41 pp.

ICES 2010. Report of the Workshop on methods for merging métiers for fishery based sampling (WKMERGE), ICES CM 2010/ACOM:40, 94 pp.

ICES 2011a. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, ICES CM 2011/ACOM:52

ICES. 2011b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2011 /ACOM:50

ICES 2011c. Report of the Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS), ICES CM 2011/ACOM:40.

ICES 2012a. Report of the second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes. ICES CM 2012/ACOM:54

ICES 2012b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2012 /ACOM:50

ICES 2012c. Report of the Working Group on Recreational Fisheries Surveys (WGRFS). ICES CM 2012/ACOM:23.

ICES 2013a. Report of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS3). ICES CM 2013/ACOM:54

ICES 2013b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2013/ACOM:56

ICES 2013c. Report of the Working Group on Recreational Fisheries Surveys. ICES CM 2013/ACOM:23.

ICES 2014a. Report of the Working Group on Commercial Catches. ICES CM 2014/ACOM:36.

ICES 2014b. Report of the Working Group on Recreational Fisheries Surveys. ICES CM 2014/ACOM:37.

ICES. 2014c. *Report of the Workshop on Developing the RDB Data Format for Design Based Sampling and Estimation (WKRDB 2014-1) 27-31 October 2014 Aberdeen, Scotland, United-Kingdom ICES WKRDB 2014-1 REPORT 2014 ICES ACOM c OMMITTEE ICES CM 2014:68*. ICES.

Jansen, Teunis, Henrik Degel, Joel Vigneau, and Ernesto Jardim. 2009. *Definition of Standard Data-Exchange Format for Sampling, Landings, and Effort Data Commercial Fisheries*. ICES COOPERATIVE RESEARCH REPORT 296. ICES.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. *https://www.R-project.org/*.

RCM NS&EA. 2014. Report of the Regional Co-ordination Meeting for the North Sea and Eastern Arctic (RCM NS&EA) 2014. Swedish University of Agriculture Sciences (SLU Aqua) Lysekil, Sweden, 08-12 September 2014

STECF 12-07 (EWG 12-01): 2012. Review of Proposed DCF 2014-2020, Part 1

Uhlmann, S, Bierman S., and van Helmond, A. (2011) A method of detecting patterns in mean lengths of samples of discarded fish, applied to the self-sampling programme of the Dutch bottom-trawl fishery. ICES Journal of Marine Science; doi:10.1093/icesjms/fsr066

Vigneau, Joel, and Stephanie Mahevas. 2007. “Detecting Sampling Outliers and Sampling Heterogeneity When Catch-at-Length Is Estimated Using the Ratio Estimator.” *ICES Journal of Marine Science: Journal Du Conseil* 64 (5): 1028–32. doi:*10.1093/icesjms/fsm077*.

Annex A – Excel reference file

Tab 'Slot'

slot_name	mandatory	definition_tab
classVersion	TRUE	slot_classVer
desc	FALSE	slot_desc
popData	FALSE	slot_popData
design	FALSE	slot_design
se	TRUE	slot_se
tr	TRUE	slot_tr
hh	FALSE	slot_hh
sl	FALSE	slot_sl
hl	FALSE	slot_hl
ca	FALSE	slot_ca

Tab 'slot_classVersion'

column_name	nullable	mandatory	pk	type_name	category
classVersion	FALSE	TRUE	FALSE	type_classVersion	text

Tab 'desc'

column_name	nullable	mandatory	pk	type_name	category
desc	FALSE	TRUE	FALSE	type_desc	text

Tab 'slot_popData'

column_nam	nullable	mandatory	pk	type_name	category
popData	FALSE	TRUE	FALSE	type_popDat	text

Tab 'slot_design'

column_name	nullable	mandatory	pk	type_name	category
design	FALSE	TRUE	FALSE	type_design	text

Tab ‘slot_history’

column_nam	nullable	mandatory	pk	type_name	category
history	FALSE	TRUE	FALSE	type_history	text

Tab ‘slot_se’

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	TRUE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
dataProv	FALSE	TRUE	FALSE	type_dataProv	text
sampCtry	FALSE	TRUE	FALSE	type_ctry	codelist
sampInst	FALSE	TRUE	FALSE	type_sampInst	text
sampMeth	FALSE	TRUE	FALSE	type_sampMeth	codelist
sampTeam	FALSE	FALSE	FALSE	type_sampTeam	text
seYear	FALSE	TRUE	FALSE	type_year	numeric
sampDate	FALSE	TRUE	FALSE	type_date	text
sampTime	TRUE	FALSE	FALSE	type_time	text
sampLoc	FALSE	TRUE	FALSE	type_harbour	text
sampLocType	FALSE	TRUE	FALSE	type_sampLocType	codelist
psuType	FALSE	TRUE	FALSE	type_psuType	codelist
design	FALSE	TRUE	FALSE	type_design	text
popData	TRUE	FALSE	FALSE	type_popData	text
sampScheme	FALSE	TRUE	FALSE	type_sampScheme	text
sampStrata	FALSE	TRUE	FALSE	type_sampStrata	text
sampTemporalUnit	FALSE	TRUE	FALSE	type_sampTemporalUnit	text
psuKey	FALSE	TRUE	FALSE	type_psuKey	text
psuId	FALSE	TRUE	FALSE	type_psuId	text
psuTotal	FALSE	TRUE	FALSE	type_psuTotal	numeric
psuSampled	FALSE	TRUE	FALSE	type_psuSampled	numeric
psuSampProb	FALSE	TRUE	FALSE	type_psuSampProb	numeric

Tab ‘slot_tr’

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	FALSE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
year	FALSE	TRUE	TRUE	type_year	numeric
proj	FALSE	TRUE	TRUE	type_proj	text
trpCode	FALSE	TRUE	TRUE	type_trpCode	text
sampType	FALSE	TRUE	FALSE	type_sampType	codelist
vslFlgCtry	FALSE	TRUE	FALSE	type_ctry	codelist
vslId	FALSE	FALSE	FALSE	type_vslId	text
vslLen	TRUE	FALSE	FALSE	type_vslLen	numeric
vslLenCat	TRUE	FALSE	FALSE	type_vslLenCat	text
vslPwr	TRUE	FALSE	FALSE	type_vslPwr	numeric
vslSize	TRUE	FALSE	FALSE	type_vslSize	numeric
vslSizeUnit	TRUE	FALSE	FALSE	type_vslSizeUnit	text
vslType	FALSE	TRUE	FALSE	type_vslType	codelist
foNum	TRUE	FALSE	FALSE	type_foNum	numeric
daysAtSea	TRUE	FALSE	FALSE	type_daysAtSea	numeric
voyageId	FALSE	FALSE	FALSE	type_voyageId	text
depLoc	TRUE	FALSE	FALSE	type_harbour	text
depDate	TRUE	FALSE	FALSE	type_date	text
depTime	TRUE	FALSE	FALSE	type_time	text
arvLoc	FALSE	TRUE	FALSE	type_harbour	text
arvDate	TRUE	FALSE	FALSE	type_date	text
arvTime	TRUE	FALSE	FALSE	type_time	text
ssuType	TRUE	FALSE	FALSE	type_ssuType	text
ssuKey	TRUE	FALSE	FALSE	type_ssuKey	text
ssuId	TRUE	FALSE	FALSE	type_ssuId	text
ssuTotal	FALSE	TRUE	FALSE	type_psuTotal	numeric
ssuSampled	FALSE	TRUE	FALSE	type_psuSampled	numeric
ssuSampProb	FALSE	TRUE	FALSE	type_psuSampProb	numeric

Tab ‘slot_hh’

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	FALSE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
year	FALSE	TRUE	TRUE	type_year	numeric
proj	FALSE	TRUE	TRUE	type_proj	text
trpCode	FALSE	TRUE	TRUE	type_trpCode	text
aggLev	TRUE	FALSE	FALSE	type_aggLev	codelist
landFrac	TRUE	FALSE	FALSE	type_landFrac	numeric
staNum	FALSE	TRUE	TRUE	type_staNum	numeric
foType	TRUE	FALSE	FALSE	type_foType	text
foKey	TRUE	FALSE	FALSE	type_fokey	text
foId	TRUE	FALSE	FALSE	type_foId	text
foVal	TRUE	FALSE	FALSE	type_foVal	codelist
catReg	FALSE	TRUE	FALSE	type_catReg	codelist
sppReg	FALSE	TRUE	FALSE	type_sppReg	codelist
foDate	FALSE	TRUE	FALSE	type_date	text
foTime	TRUE	FALSE	FALSE	type_time	text
foDur	TRUE	FALSE	FALSE	type_foDur	numeric
latIni	TRUE	FALSE	FALSE	type_lat	numeric
lonIni	TRUE	FALSE	FALSE	type_lon	numeric
latFin	TRUE	FALSE	FALSE	type_lat	numeric
lonFin	TRUE	FALSE	FALSE	type_lon	numeric
ecoZone	TRUE	FALSE	FALSE	type_ecozone	text
area	FALSE	TRUE	FALSE	type_area	text
rect	TRUE	FALSE	FALSE	type_rect	text
subRect	TRUE	FALSE	FALSE	type_subRect	text
foDep	TRUE	FALSE	FALSE	type_foDep	numeric
waterDep	TRUE	FALSE	FALSE	type_waterDep	numeric
foCatNat	TRUE	FALSE	FALSE	type_foCatNat	text
foCatEu5	TRUE	FALSE	FALSE	type_foCatEu5	codelist
foCatEu6	FALSE	TRUE	FALSE	type_foCatEu6	codelist
gear	FALSE	TRUE	FALSE	type_gearType	codelist
meshSize	TRUE	FALSE	FALSE	type_meshSize	numeric
selDev	TRUE	FALSE	FALSE	type_selDev	text
meshSizeSelDev	TRUE	FALSE	FALSE	type_meshSizeSelDev	numeric
landCtry	FALSE	TRUE	FALSE	type_ctry	codelist
landLoc	TRUE	FALSE	FALSE	type_harbour	text
landlocType	TRUE	FALSE	FALSE	type_landlocType	text
landDate	TRUE	FALSE	FALSE	type_date	text
landTime	TRUE	FALSE	FALSE	type_time	text
saleCtry	TRUE	FALSE	FALSE	type_ctry	codelist
saleLoc	TRUE	FALSE	FALSE	type_harbour	text
saleDate	TRUE	FALSE	FALSE	type_date	text
saleTime	TRUE	FALSE	FALSE	type_time	text
buyerLoc	TRUE	FALSE	FALSE	type_harbour	text
domain1	TRUE	FALSE	FALSE	type_domain1	text
domain2	TRUE	FALSE	FALSE	type_domain2	text
foTotal	TRUE	FALSE	FALSE	type_foNum	numeric
foSampled	TRUE	FALSE	FALSE	type_foNum	numeric
foSampProb	TRUE	FALSE	FALSE	type_psuSampProb	numeric

Tab ‘slot_sl’

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	FALSE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
year	FALSE	TRUE	TRUE	type_year	numeric
proj	FALSE	TRUE	TRUE	type_proj	text
trpCode	FALSE	TRUE	TRUE	type_trpCode	text
staNum	FALSE	TRUE	TRUE	type_staNum	numeric
foId	TRUE	FALSE	TRUE	type_foId	text
commSpp	FALSE	TRUE	FALSE	type_commSpp	text
spp	FALSE	TRUE	TRUE	type_spp	text
catchCat	FALSE	TRUE	TRUE	type_catchCat	codelist
landCat	FALSE	TRUE	TRUE	type_landCat	codelist
commCatScl	TRUE	FALSE	TRUE	type_commCatScl	codelist
commCat	TRUE	FALSE	TRUE	type_commCat	codelist
subSampCat	TRUE	FALSE	TRUE	type_subSampCat	codelist
sex	TRUE	FALSE	TRUE	type_sex	codelist
unitType	TRUE	FALSE	FALSE	type_unitType	text
unitKey	TRUE	FALSE	FALSE	type_unitKey	text
unitId	TRUE	FALSE	TRUE	type_unitIf	text
wt	FALSE	TRUE	FALSE	type_wt	numeric
subSampWt	TRUE	FALSE	FALSE	type_subSampWt	numeric
totWtDeriv	TRUE	FALSE	FALSE	type_wt	numeric
sampWtDeriv	TRUE	FALSE	FALSE	type_subSampWt	numeric
measType	TRUE	FALSE	FALSE	type_measType	text
pres	TRUE	FALSE	FALSE	type_pres	text
convFacWt	FALSE	FALSE	FALSE	type_convFacWt	numeric
lenCode	FALSE	FALSE	FALSE	type_lenCode	codelist
unitTotal	TRUE	FALSE	FALSE	type_unitTotal	numeric
unitSampled	TRUE	FALSE	FALSE	type_unitSampled	numeric
unitSampProb	TRUE	FALSE	FALSE	type_psuSampProb	numeric

Tab 'slot_hl'

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	FALSE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
year	FALSE	TRUE	TRUE	type_year	numeric
proj	FALSE	TRUE	TRUE	type_proj	text
trpCode	FALSE	TRUE	TRUE	type_trpCode	text
staNum	FALSE	TRUE	TRUE	type_staNum	numeric
foId	TRUE	FALSE	TRUE	type_foId	text
spp	FALSE	TRUE	TRUE	type_spp	text
catchCat	FALSE	TRUE	TRUE	type_catchCat	codelist
landCat	FALSE	TRUE	TRUE	type_landCat	codelist
commCatScl	TRUE	TRUE	TRUE	type_commCatScl	codelist
commCat	TRUE	TRUE	TRUE	type_commCat	codelist
subSampCat	TRUE	TRUE	TRUE	type_subSampCat	codelist
sex	TRUE	TRUE	TRUE	type_sex	codelist
unitId	TRUE	FALSE	TRUE	type_unitIf	text
indSex	TRUE	TRUE	FALSE	type_sex	codelist
lenCls	FALSE	TRUE	TRUE	type_lenCls	numeric
lenNum	FALSE	TRUE	FALSE	type_lenNum	numeric
measType	FALSE	FALSE	FALSE	type_measType	text
measCls	FALSE	FALSE	FALSE	type_measCls	text
measNum	FALSE	FALSE	FALSE	type_measNum	numeric
convFacLen	TRUE	FALSE	FALSE	type_convFacLen	numeric
fishTotal	FALSE	FALSE	FALSE	type_fishTotal	numeric
fishSampled	FALSE	FALSE	FALSE	type_fishSampled	numeric
fishSampProb	FALSE	FALSE	FALSE	type_fishSampProb	numeric

Tab 'slot_ca'

column_name	nullable	mandatory	pk	type_name	category
recType	FALSE	TRUE	FALSE	type_recType	codelist
seCode	FALSE	TRUE	TRUE	type_seCode	text
year	TRUE	FALSE	TRUE	type_year	numeric
proj	TRUE	FALSE	TRUE	type_proj	text
trpCode	TRUE	FALSE	TRUE	type_trpCode	text
staNum	FALSE	TRUE	FALSE	type_staNum	numeric
foId	TRUE	FALSE	FALSE	type_foId	text
quarter	TRUE	FALSE	FALSE	type_quarter	numeric
month	TRUE	FALSE	FALSE	type_month	numeric
spp	TRUE	FALSE	FALSE	type_spp	text
sex	TRUE	FALSE	FALSE	type_sex	codelist
unitId	TRUE	FALSE	FALSE	type_unitIf	text
indSex	TRUE	TRUE	FALSE	type_sex	codelist
catchCat	TRUE	FALSE	FALSE	type_catchCat	codelist
landCat	TRUE	FALSE	FALSE	type_landCat	codelist
commCatScl	TRUE	FALSE	FALSE	type_commCatScl	codelist
commCat	TRUE	FALSE	FALSE	type_commCat	codelist
subSampCat	TRUE	FALSE	FALSE	type_subSampCat	text
stock	TRUE	FALSE	FALSE	type_stock	text
area	TRUE	FALSE	FALSE	type_area	text
rect	TRUE	FALSE	FALSE	type_rect	text
subRect	TRUE	FALSE	FALSE	type_subRect	text
lenCls	TRUE	FALSE	FALSE	type_lenCls	numeric
age	TRUE	FALSE	FALSE	type_age	numeric
fishId	TRUE	FALSE	FALSE	type_fishId	text
lenCode	TRUE	FALSE	FALSE	type_lenCode	codelist
measType	FALSE	FALSE	FALSE	type_measType	text
measCls	FALSE	FALSE	FALSE	type_measCls	text
fishAtLengthTotal	FALSE	FALSE	FALSE	type_fishAtLengthTotal	numeric
fishAtlengthSampled	FALSE	FALSE	FALSE	type_fishAtlengthSampled	numeric
individualFishSampPr	FALSE	FALSE	FALSE	type_individualFishSampProb	numeric
ageMeth	TRUE	FALSE	FALSE	type_ageMeth	codelist
plusGrp	TRUE	FALSE	FALSE	type_plusGrp	codelist
otoWt	TRUE	FALSE	FALSE	type_otoWt	numeric
otoSide	TRUE	FALSE	FALSE	type_otoSide	codelist
indWt	TRUE	FALSE	FALSE	type_indWt	numeric
matMeth	TRUE	FALSE	FALSE	type_matMeth	codelist
matScale	TRUE	FALSE	FALSE	type_matScale	codelist
matStage	TRUE	FALSE	FALSE	type_matStage	codelist

Tab 'date_types'

type_name

Tab 'text_types'

type_name
type_classVersion
type_desc
type_popData
type_design
type_history
type_seCode
type_dataProv
type_sampInst
type_sampTeam
type_date
type_time
type_harbour
type_design
type_popData
type_sampScheme
type_sampStrata
type_sampTempora
type_psuKey
type_psuId
type_proj
type_trpCode
type_vslId
type_vslLenCat
type_vslSizeUnit
type_voyageId
type_ssuType
type_ssuKey
type_ssuId
type_foType
type_fokey
type_foId
type_ecozone
type_area
type_rect
type_subRect
type_foCatNat
type_selDev
type_landlocType
type_domain1
type_domain2
type_spp
type_commSpp
type_unitType
type_unitKey
type_unitIf
type_pres
type_measType
type_measType
type_measCls
type_stock
type_subSampCat

Tab ‘numeric_types’

type_name	is_integer	min	max
type_year	TRUE	1900	2020
type_psuTotal	TRUE	0	10000000
type_psuSampled	TRUE	0	2000
type_psuSampProb	FALSE	0	1
type_vslLen	TRUE	3	160
type_vslPwr	TRUE	4	8500
type_vslSize	TRUE	1	2500
type_foNum	TRUE	1	300
type_daysAtSea	TRUE	1	90
type_landFrac	FALSE	0	1
type_staNum	TRUE	0	1.00E+015
type_foDur	TRUE	0	86400
type_lat	FALSE	-90	90
type_lon	FALSE	-180	180
type_foDep	TRUE	0	999
type_waterDep	TRUE	0	9999
type_meshSize	TRUE	1	999
type_meshSizeSelDev	TRUE	18	200
type_wt	FALSE	0	1.00E+015
type_subSampWt	FALSE	0	1.00E+015
type_convFacWt	FALSE	0	1.00E+009
type_unitTotal	TRUE	1	1.00E+015
type_unitSampled	TRUE	1	1.00E+015
type_lenNum	FALSE	1	999
type_lenCls	FALSE	1	999
type_quarter	TRUE	1	4
type_month	TRUE	1	12
type_age	TRUE	0	99
type_otoWt	FALSE	0	100
type_indWt	FALSE	1	100000
type_measNum	TRUE		
type_convFacLen	FALSE		
type_fishTotal	TRUE		
type_fishSampled	TRUE		
type_fishSampProb	FALSE	0	1
type_fishAtLengthTotal	TRUE		
type_fishAtlengthSampled	TRUE		
type_individualFishSampProb	FALSE	0	1

Tab ‘logical_types’

type_name

Tab ‘codelist_types’

type_name	enumeration_table
type_recType	codelist_recType
type_ctry	codelist_ctry
type_sampMeth	codelist_sampMeth
type_sampLocType	codelist_sampLocType
type_psuType	codelist_psuType
type_sampType	codelist_sampType
type_vslType	codelist_vslType
type_aggLev	codelist_aggLev
type_foVal	codelist_foVal
type_catReg	codelist_catReg
type_sppReg	codelist_sppReg
type_foCatEu5	codelist_foCatEu5
type_foCatEu6	codelist_foCatEu6
type_gearType	codelist_gearType
type_catchCat	codelist_catchCat
type_landCat	codelist_landCat
type_lenCode	codelist_lenCode
type_sampMeth	codelist_sampMeth
type_foVal	codelist_foVal
type_commCatScl	codelist_commCatScl
type_commCat	codelist_commCat
type_subSampCat	codelist_subSampCat
type_sex	codelist_sex
type_ageMeth	codelist_ageMeth
type_plusGrp	codelist_plusGrp
type_otoSide	codelist_otoSide
type_matMeth	codelist_matMeth
type_matScale	codelist_matScale
type_matStage	codelist_matStage

Tab ‘codelist_recType’

CODE	DESCRIPTION
se	sampling event
tr	trip
hh	haul
sl	species list
hl	length
ca	biological parameters

Tab ‘codelist_ctry’

CODE	DESCRIPTION
ABW	Aruba
AFG	Afghanistan
AGO	Angola
AIA	Anguilla
ALA	Åland Islands
ALB	Albania
AND	Andorra
ARE	United Arab Emirates
ARG	Argentina
ARM	Armenia
ASM	American Samoa
ATA	Antarctica
ATF	French Southern Territories
ATG	Antigua and Barbuda
AUS	Australia
AUT	Austria
AZE	Azerbaijan
BDI	Burundi
BEL	Belgium
BEN	Benin
BES	Bonaire, Sint Eustatius and Saba
BFA	Burkina Faso
BGD	Bangladesh
BGR	Bulgaria
BHR	Bahrain
BHS	Bahamas
BIH	Bosnia and Herzegovina
BLM	Saint Barthélemy
BLR	Belarus
BLZ	Belize
BMU	Bermuda
BOL	Bolivia, Plurinational State of
BRA	Brazil
BRB	Barbados
BRN	Brunei Darussalam
BTN	Bhutan
BVT	Bouvet Island
BWA	Botswana
CAF	Central African Republic
CAN	Canada
CCK	Cocos (Keeling) Islands
CHE	Switzerland
CHL	Chile
CHN	China
CIV	Côte d'Ivoire
CMR	Cameroon
COD	Congo, the Democratic Republic of the
COG	Congo
COK	Cook Islands
COL	Colombia
COM	Comoros
CPV	Cabo Verde
CRI	Costa Rica
CUB	Cuba
CUW	Curaçao
CXR	Christmas Island
CYM	Cayman Islands
CYP	Cyprus
CZE	Czech Republic
DEU	Germany
DJI	Djibouti
DMA	Dominica
DNK	Denmark
DOM	Dominican Republic
DZA	Algeria
ECU	Ecuador
EGY	Egypt
ERI	Eritrea
ESH	Western Sahara
ESP	Spain
EST	Estonia
ETH	Ethiopia
FIN	Finland
FJI	Fiji
FLK	Falkland Islands (Malvinas)
FRA	France
FRO	Faroe Islands
FSM	Micronesia, Federated States of
GAB	Gabon
GBR	United Kingdom
GEO	Georgia
GGY	Guernsey
GHA	Ghana
GIB	Gibraltar
GIN	Guinea
GLP	Guadeloupe
GMB	Gambia
GNB	Guinea-Bissau
GNQ	Equatorial Guinea
GRC	Greece
GRD	Grenada
GRL	Greenland
GTM	Guatemala
GUF	French Guiana
GUM	Guam
GUY	Guyana
HKG	Hong Kong
HMD	Heard Island and McDonald Islands
HND	Honduras
HRV	Croatia
HTI	Haiti
HUN	Hungary
IDN	Indonesia
IMN	Isle of Man
IND	India
IOT	British Indian Ocean Territory
IRL	Ireland
IRN	Iran, Islamic Republic of
IRQ	Iraq
ISL	Iceland
ISR	Israel
ITA	Italy
JAM	Jamaica
JEY	Jersey
JOR	Jordan
JPN	Japan
KAZ	Kazakhstan
KEN	Kenya
KGZ	Kyrgyzstan
KHM	Cambodia
KIR	Kiribati
KNA	Saint Kitts and Nevis
KOR	Korea, Republic of
KWT	Kuwait
LAO	Lao People's Democratic Republic
LBN	Lebanon
LBR	Liberia
LBY	Libya
LCA	Saint Lucia
LIE	Liechtenstein
LKA	Sri Lanka
LSO	Lesotho
LTU	Lithuania
LUX	Luxembourg
LVA	Latvia
MAC	Macao
MAF	Saint Martin (French part)
MAR	Morocco
MCO	Monaco
MDA	Moldova, Republic of
MDG	Madagascar
MDV	Maldives
MEX	Mexico
MHL	Marshall Islands
MKD	Macedonia, the former Yugoslav Reput
MLI	Mali
MLT	Malta
MMR	Myanmar
MNE	Montenegro
MNG	Mongolia
MNP	Northern Mariana Islands
MOZ	Mozambique
MRT	Mauritania
MSR	Montserrat
MTQ	Martinique
MUS	Mauritius
MWI	Malawi
MYS	Malaysia
MYT	Mayotte
NAM	Namibia
NCL	New Caledonia
NER	Niger
NFK	Norfolk Island
NGA	Nigeria
NIC	Nicaragua
NIU	Niue
NLD	Netherlands
NOR	Norway
NPL	Nepal
NRU	Nauru
NZL	New Zealand
OMN	Oman
PAK	Pakistan
PAN	Panama
PCN	Pitcairn
PER	Peru
PHL	Philippines
PLW	Palau
PNG	Papua New Guinea
POL	Poland
PRI	Puerto Rico
PRK	Korea, Democratic People's Republic of
PRT	Portugal
PRY	Paraguay
PSE	Palestine, State of
PYF	French Polynesia
QAT	Qatar
REU	Réunion
ROU	Romania
RUS	Russian Federation
RWA	Rwanda
SAU	Saudi Arabia
SDN	Sudan
SEN	Senegal
SGP	Singapore
SGS	South Georgia and the South Sandwich Islands
SHN	Saint Helena, Ascension and Tristan da Cunha
SJM	Svalbard and Jan Mayen
SLB	Solomon Islands
SLE	Sierra Leone
SLV	El Salvador
SMR	San Marino
SOM	Somalia
SPM	Saint Pierre and Miquelon
SRB	Serbia
SSD	South Sudan
STP	Sao Tome and Principe
SUR	Suriname
SVK	Slovakia
SVN	Slovenia
SWE	Sweden
SWZ	Swaziland
SXM	Sint Maarten (Dutch part)
SYC	Seychelles
SYR	Syrian Arab Republic
TCA	Turks and Caicos Islands
TCD	Chad
TGO	Togo
THA	Thailand
TJK	Tajikistan
TKL	Tokelau
TKM	Turkmenistan
TLS	Timor-Leste
TON	Tonga
TTO	Trinidad and Tobago
TUN	Tunisia
TUR	Turkey
TUV	Tuvalu
TWN	Taiwan, Province of China
TZA	Tanzania, United Republic of
UGA	Uganda
UKR	Ukraine
UMI	United States Minor Outlying Islands
URY	Uruguay
USA	United States of America
UZB	Uzbekistan
VAT	Holy See (Vatican City State)
VCT	Saint Vincent and the Grenadines
VEN	Venezuela, Bolivarian Republic of
VGB	Virgin Islands, British
VIR	Virgin Islands, U.S.
VNM	Viet Nam
VUT	Vanuatu
WLF	Wallis and Futuna
WSM	Samoa
YEM	Yemen
ZAF	South Africa
ZMB	Zambia
ZWE	Zimbabwe

Tab ‘codelist_sampMeth’

CODE	DESCRIPTION
Observer	Sampling by observers
SelfSampling	Sampling by fishers

Tab ‘codelist_sampLocType’

CODE	DESCRIPTION
market	market sampling
port	port sampling
vessel	vessel sampling

Tab ‘codelist_psuType’

CODE	DESCRIPTION
vessel_time	
trip	
site_day	
site	

Tab ‘codelist_sampType’

CODE	DESCRIPTION
S	sea sampling
M	market sampling of known fishing trips
D	market sampling of mixed trips
V	vendor

Tab ‘codelist_vslType’

CODE	DESCRIPTION
1	Stem trawler
2	Sea side trawler
3	Gill-netter
4	Other boats

Tab ‘codelist_foVal’

CODE	DESCRIPTION
I	Invalid fishing operation
V	Valid fishing operation

Tab ‘codelist_aggLev’

CODE	DESCRIPTION
H	Haul
T	Trip

Tab ‘codelist_catReg’

CODE	DESCRIPTION
ALL	The part of the catch that was registered. SL record is expected for both landings and discards fractions.
LAN	SL record is expected only for the landed fraction.
DIS	SL record is expected only for the discarded fraction.
NON	There are no SL records. Species registration must also be assigned “Non”

Tab ‘codelist_sppReg’

CODE	DESCRIPTION
ALL	SL record is expected for all species in the given part of the catch
PAR	Partial: SL record is expected only for some of the caught species
NON	None: There are no SL records.

Tab ‘codelist _foCatEu5’

CODE	DESCRIPTION

Tab ‘codelist_foCatEu6’

CODE
DESCRIPTION

Active
All
Belgium
BOT
Bottom trawl
BT-Fi-Bal
C-Allgears
C-Beam
C-Fixed
C-other
C-Otter
C-Purse
Danish seine
DemIND
DE-PTF
DK-ALL
DK-IND
DN-her
DRB_all_0_0_all
DRB_MOL_0_0_0_all
ENGALL
EN-HL
EN-PFT
ES-ALL
FA-PS
Fleet-A
Fleet-All
Fleet-B
Fleet-C
Fleet-D
Fleet-F
Fleet-ind
FPO_CRU_0_0_0_all
FPO_DEF
FPO_MOL_0_0_0_all
FRAALL
FRATRB_IV
FR-PTF
G.._DEF<10-18_0
GERALL
GIL
Gillnet
GL-ALL
GN
GN_DEF_100-109_0_0_all
GNS
GNS_CRU_0_0_0_all
GNS_DEF
GNS_DEF_>=100_0_0
GNS_DEF_>=220_0_0_all
GNS_DEF_>=220_0_0_all_FDF
GNS_DEF_100-119_0_0_all
GNS_DEF_100-119_0_0_all_FDF
GNS_DEF_10-30_0_0_all
GNS_DEF_120-219_0_0
GNS_DEF_120-219_0_0_all
GNS_DEF_120-219_0_0_all_FDF
GNS_DEF_60-79_0_0
GNS_DEF_80-99_0_0
GNS_DEF_90-99_0_0_all
GNS_DEF_all_0_0_all
GNS_DWS
GNS_SPF_50-59_0_0_all
GNS-DEF
GTR_CRU_0_0_0_all
GTR_DEF
GTR_DEF_>=220_0_0_all
GTR_DEF_100-119_0_0_all
GTR_DEF_120-219_0_0_all
GTR_DEF_60-79_0_0
GTR_DEF_90-99_0_0_all
GTR_DEF_all_0_0_all
GY-ALL
Handline
hernoss-fleet
HM-ALL
IC-ALL
Ice_herring_fleet
INDspr
Inshore-Gillnets
Inshore-trawlers
IOM-ALL
IR-PTF
IS_GNS
IS_OTB
IS_OTHER
IS_OTM
IS_SSC
IS_TBB
IS_UNKNOWN
IS-ALL
JY-ALL
LHM_DEF
LHM_DEF_0_0_0
LHM_DWS
LHM_DWS_0_0_0
Lines and Jigging
LLS_DEF
LLS_DEF_0_0_0
LLS_DEF_0_0_0_all
LLS_DWS
LLS_FIF_0_0_0_all
LLS_FIF_0_0_0_all_FDF
LLS-DEF
Longline
LT-ALL
Mackerel ByCatch
Mackerel Discards
Mackerel Pelagic Trawl
Mackerel Purse Seine
MIF
MIS
MIS_DEF_all_0_0_all
MIS_MIS
MIS_MIS_0_0_0
MIS_MIS_0_0_0_HC
MIS_MIS_0_0_0_HC_FDF
MIS_MIS_0_0_0_IBC
MIS_MIS_All_0_0_All
MIS_MIS_trial_run
nep-11
nep-12
nep-13
Nephrops creels in FU3
Nephrops trawls in FU3
Nephrops trawls in FU4
NI-PT
NL-PTF
NO-HL
NO-PS
Offshore-Gillnets
Offshore-trawlers
OTB
OTB- DWS
OTB_CRU
OTB_CRU_>=70_0_0
OTB_CRU_100-119_0_0_all
OTB_CRU_100-119_1_0_all
OTB_CRU_16-31_0_0_all
OTB_CRU_32-69_0_0_all
OTB_CRU_32-69_2_22_all
OTB_CRU_70-89_2_35_all
OTB_CRU_70-99_0_0
OTB_CRU_70-99_0_0_all
OTB_CRU_70-99_0_0_all_FDF
OTB_CRU_70-99_1_0_all
OTB_CRU_70-99_1_110_all
OTB_CRU_70-99_2_0_all
OTB_CRU_90-119_0_0_all
OTB_CRU_90-119_0_0_all_FDF
OTB_CRU_All_0_0_All
OTB_CRU_IE
OTB_DEF
OTB_DEF_<16_0_0_all
OTB_DEF_>=120_0_0
OTB_DEF_>=120_0_0_<24
OTB_DEF_>=120_0_0_>=40
OTB_DEF_>=120_0_0_24<40
OTB_DEF_>=120_0_0_all
OTB_DEF_>=120_0_0_all_FDF
OTB_DEF_>=55_0_0
OTB_DEF_>=70_0_0
OTB_DEF_>40_>=100_POK
OTB_DEF_100-119_0_0
OTB_DEF_100-119_0_0_>=40m__POK
OTB_DEF_100-119_0_0_all
OTB_DEF_100-119_0_0_all_FDF
OTB_DEF_100-119_1_100_all
OTB_DEF_70-99_0_0
OTB_DEF_70-99_0_0_<24
OTB_DEF_70-99_0_0_all
OTB_DEF_70-99_0_0_all_FDF
OTB_DEF_80-99_0_0
OTB_DEF_All_0_0_All
OTB_DWS
OTB_DWS_>=120_0_0_all
OTB_DWS_100-119_0_0_all
OTB_MCD_>=55_0_0
OTB_MCD_70-99_0_0_all
OTB_MCF_>=70_0_0
OTB_MOL_>=120_0_0_all
OTB_MOL_100-119_0_0_all
OTB_MOL_70-99_0_0_all
OTB_MPD_>=55_0_0
OTB_SPF_32-69_0_0_all
OTB_SPF_70-99_0_0_all
OTB-CRUST
OTB-DEF
Oth
Other
Other_unsp
OTM_DEF
OTM_DEF_100-119_0_0_all
OTM_DEF_32-54_0_0_all
OTM_DEF_32-69_0_0_all
OTM_DEF_55-69_0_0_all
OTM_DEF_70-99_0_0_all
OTM_DEF_80-89_0_0_all
OTM_DWS
OTM_MPD
OTM_SPF_16-31_0_0
OTM_SPF_32-69_0_0_all
OTM1-her
OTM2-her
OTM-her
OTM-her-vian
OTT_CRU_>=70_0_0
OTT_CRU_100-119_0_0
OTT_DEF_>=120_0_0_all
OTT_DEF_>=70_0_0
OTT_DEF_100-119_0_0
OTT_DEF_100-119_0_0_all
OTT_DEF_16-31_0_0_all
OTT_DEF_80-89_0_0_all
OTT_DEF_90-99_0_0_all
OTT-CRU
OTT-DEF
OTT-DWS
Passive
Passive gears
PEL
Pel Freezer Trawler
Pelagic trawl
Pelagic trawlers
PL-ALL
PORART
PORBT
PS_DEF_32-54_0_0
PS_SPF_0_0_0
PS-ANE
PS-PIL
PT_ALL
PT-ALL
PTB_all_0_0_all
PTB_DEF_>=120_0_0_all
PTB_DEF_>=70_0_0
PTB_DEF_70-99_0_0_all
PTB_DEF_80-89_0_0_all
PTB_DEF_all_0_0_all
PTB_MPD
PTB_MPD_>=55_0_0
PTM_DEF_100-119_0_0_all
PTM_DEF_90-104_0_0
PTM_SPF_32-69_0_0_all
PTM-DEF
PTM-her
PTM-her-nirs
Purse seine
PV-ANE
RING_her
RU-PTF
SC-ALL
SC-DIS
SCO_4Gears
SCOLTR
SCONTR
SCOOTH
SCOPTR
SCOSEI
SCOTRL
SDN_all_0_0_all
SDN_DEF
SDN_DEF_>=120_0_0_all
SDN_DEF_>=120_0_0_all_FDF
SDN_DEF_70-99_0_0_all
SDN_SPF
SP_ALL
SPABT
SPAGIL
SSC
SSC_DEF
SSC_DEF_>=120_0_0_all
SSC_DEF_>=120_0_0_all_FDF
SSC_DEF_100-119_0_0_all
SSC_DEF_100-119_0_0_all_FDF
SSC_DEF_70-99_0_0_all
SSC_DEF_70-99_0_0_all_FDF
SSC_DEF_80-89_0_0_all
SSC_DEF_All_0_0_All
SSC-DEF
SW-ALL
TBB
TBB_CRU_16-31_0_0_all
TBB_DEF
TBB_DEF_<16_0_0_all
TBB_DEF_>=120_0_0_all
TBB_DEF_100-119_0_0_all
TBB_DEF_100-119_0_0_all_FDF
TBB_DEF_70-99_0_0_all
TBB_DEF_70-99_0_0_all_FDF
TBB_DEF_90-99_0_0_all
TBB_DEF_all_0_0_all
TBB-DEF
TestA
TestB
TestC
TestD
Trapnet
Trawl
TrawlPlus
U10
v1s1t1
v1s1t2
v2s1t1
v2s1t2
v6s1t1
v6s1t2
Winter Seine

Tab 'codelist_gearType'

CODE	DESCRIPTION		
PS	Surrounding nets - With purse lines (purse seines)	GNS	Gillnets and entangling nets - Set gillnets (anchored)
LA	Surrounding nets - Without purse lines (lampara)	GND	Gillnets and entangling nets - Driftnets
SB	Seine nets - Beach seines	GNC	Gillnets and entangling nets - Encircling gillnets
SV	Seine nets - Boat or vessel seines	GNF	Gillnets and entangling nets - Fixed gillnets (on stakes)
SDN	Seine nets - Danish seines	GTR	Gillnets and entangling nets - Trammel nets
SSC	Seine nets - Scottish seines	GTN	Gillnets and entangling nets - Combined gillnets – trammel nets
SPR	Seine nets - Pair seines	GEN	Gillnets and entangling nets - Gillnets and entangling nets (not specified)
SX	Seine nets - Seine nets (not specified)	GN	Gillnets and entangling nets - Gillnets (not specified)
TBB	Trawls - Bottom beam trawls	FPN	Traps - Stationary uncovered pound nets
OTB	Trawls - Bottom otter trawls	FPO	Traps - Pots
PTB	Trawls - Bottom pair trawls	FYK	Traps - Fyke nets
TBN	Trawls - Bottom Nephrops trawls	FSN	Traps - Stow nets
TBS	Trawls - Bottom shrimp trawls	FWR	Traps - Barriers, fences, weirs, etc.
TB	Trawls - Bottom trawls (not specified)	FAR	Traps - Aerial traps
OTM	Trawls - Midwater otter trawls	FIX	Traps - Traps (not specified)
PTM	Trawls - Midwater pair trawls	LHP	Hooks and Lines - Handlines and pole-lines (hand operated)
TMS	Trawls - Midwater shrimp trawls	LHM	Hooks and Lines - Handlines and pole-lines (mechanized)
OTT	Trawls - Otter twin trawls	LLS	Hooks and Lines - Set longlines
OT	Trawls - Otter trawls (not specified)	LLD	Hooks and Lines - Drifting longlines
PT	Trawls - Pair trawls (not specified)	LL	Hooks and Lines - Longlines (not specified)
TX	Trawls - Other trawls (not specified)	LTL	Hooks and Lines - Trolling lines
DRB	Dredges - Boat dredges	LX	Hooks and Lines - Hooks and lines (not specified)
DRH	Dredges - Hand dredges	HAR	Grappling and wounding - Harpoons
LNP	Lift nets - Portable lift nets	HMP	Harvesting machines - Pumps
LNB	Lift nets - Boat-operated lift nets	HMD	Harvesting machines - Mechanized dredges
LNS	Lift nets - Shore-operated stationary lift nets	HMX	Harvesting machines - Harvesting machines (not specified)
LN	Lift nets - Lift nets (not specified)	MIS	Miscellaneous gear
FCN	Falling gear - Cast nets	RG	Recreational fishing gear
FG	Falling gear - Falling gear (not specified)	NK	Gear not know or not specified

Tab 'codelist_catchCat'

CODE	DESCRIPTION
DIS	Discard
LAN	Landing

Tab 'codelist_landCat'

CODE	DESCRIPTION
IND	Industry
HUC	Human consumption

Tab 'codelist_catCommScale'

CODE	DESCRIPTION
UNSORTED	Discard

Tab ‘codelist_commCat’

CODE	DESCRIPTION
1	1
2	2
3	3
4	4
5	5

Tab ‘codelist_subSampCat’

CODE	DESCRIPTION
SMALL	Small
MEDIUM	Medium
LARGE	Large

Tab ‘codelist_sex’

CODE	DESCRIPTION
F	Female
M	Male
T	Transitional (during sex change)
I	Indetermined
N	Not determined

Tab ‘codelist_lenCode’

CODE	DESCRIPTION
MM	mm
SCM	Semi-centimetre (0.5 cm)
CM	cm
25 MM	25 mm
5 CM	5 cm

Tab ‘codelist_ageMeth’

CODE	DESCRIPTION
OWR	Otolith (winter rings)
SCALE	Scale
VERTEBRA	Vertebra
DORSAL SPINE	Dorsal spine
ANAL SPINE	Anal spine
ILICIUM	Ilicium

Tab ‘codelist_plusGrp’

CODE	DESCRIPTION
-	Not an “age-plus-group”
+	An “age-plus-group”

Tab ‘codelist_otoSide’

CODE	DESCRIPTION
R	Right
L	Left

Tab ‘codelist_matMeth’

CODE	DESCRIPTION
MACR	Macroscopical
HIST	Histological

Tab ‘codelist_matScale’

CODE	DESCRIPTION
MEDITS	Medits

Tab ‘codelist_matStage’

CODE	DESCRIPTION
0	0
1	1
2	2
2A	2A
2B	2B
2C	2C
2D	2D
2E	2E
3	3
3A	3A
3B	3B
3C	3C
3D	3D
4	4
4A	4A
4B	4B
5	5

Annex B - Output of the quality checks

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	correctionFlag	T	Sampling_ty	Landing_count	Vessel_flag_countr	Year	Project	Trip_numb	Vessel_lengt	Vessel_pow	Vessel_siz	Vessel_typ	Harb	No_SetsHauls_on_tr	Days_at_s	Vessel_identifier	Sampling_coun	Sampling_method
761		TR	M	FRA	FRA	2015	SIH-OBSVENTE	24011696	20	316.0	72		XST			642405	FRA	37.07-OTB_DEF - Observer - TCC
762		TR	M	FRA	FRA	2015	SIH-OBSVENTE	24011694	24	316.0	100		XST			923734	FRA	37.07-OTB_DEF - Observer - TCC
763		TR	M	FRA	FRA	2015	SIH-OBSVENTE	24011693	20	316.0	40		XST			670404	FRA	37.07-OTB_DEF - Observer - TCC
764		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981098	24	316.0	134		XST			900282	FRA	37.07-OTB_DEF - Observer - TCC
765		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981097	26	316.0	71		XST			374056	FRA	37.07-OTB_DEF - Observer - TCC
766	no fish for this trip/	TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981096	24	316.0	134		XST			900283	FRA	37.07-OTB_DEF - Observer
767		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981095	24	316.0	80		XST			595980	FRA	37.07-OTB_DEF - Observer - TCC
768		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981093	20	316.0	40		XST			670404	FRA	37.07-OTB_DEF - Observer - TCC
769		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981092	24	316.0	80		XST			436675	FRA	37.07-OTB_DEF - Observer - TCC
770		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981091	24	316.0	49		XST			330140	FRA	37.07-OTB_DEF - Observer - TCC
771		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981090	24	316.0	79		XST			613327	FRA	37.07-OTB_DEF - Observer - TCC
772		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981089	24	316.0	134		XST			900283	FRA	37.07-OTB_DEF - Observer - TCC
773		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981088	24	316.0	80		XST			595980	FRA	37.07-OTB_DEF - Observer - TCC
774		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981086	24	316.0	134		XST			900282	FRA	37.07-OTB_DEF - Observer - TCC
775		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981085	24	316.0	80		XST			436675	FRA	37.07-OTB_DEF - Observer - TCC
776		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981084	26	316.0	71		XST			374056	FRA	37.07-OTB_DEF - Observer - TCC
777		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981083	20	316.0	40		XST			670404	FRA	37.07-OTB_DEF - Observer - TCC
778		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981082	24	316.0	49		XST			330140	FRA	37.07-OTB_DEF - Observer - TCC
779		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981081	24	316.0	76		XST			624710	FRA	37.07-OTB_DEF - Observer - TCC
780		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23981080	24	316.0	79		XST			613327	FRA	37.07-OTB_DEF - Observer - TCC
781		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935686	24	316.0	100		XST			923734	FRA	37.07-OTB_DEF - Observer - TCC
782		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935684	24	316.0	134		XST			900283	FRA	37.07-OTB_DEF - Observer - TCC
783		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935683	24	316.0	80		XST			436675	FRA	37.07-OTB_DEF - Observer - TCC
784		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935682	24	316.0	134		XST			900282	FRA	37.07-OTB_DEF - Observer - TCC
785		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935681	20	316.0	40		XST			670404	FRA	37.07-OTB_DEF - Observer - TCC
786		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935680	24	316.0	49		XST			330140	FRA	37.07-OTB_DEF - Observer - TCC
787		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23935679	24	316.0	79		XST			613327	FRA	37.07-OTB_DEF - Observer - TCC
788		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934808	24	316.0	99		XST			923715	FRA	37.07-OTB_DEF - Observer - TCC
789		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934807	26	316.0	71		XST			374056	FRA	37.07-OTB_DEF - Observer - TCC
790		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934806	24	316.0	134		XST			900282	FRA	37.07-OTB_DEF - Observer - TCC
791		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934805	24	316.0	100		XST			923734	FRA	37.07-OTB_DEF - Observer - TCC
792		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934804	24	316.0	134		XST			900283	FRA	37.07-OTB_DEF - Observer - TCC
793		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934803	24	316.0	80		XST			595980	FRA	37.07-OTB_DEF - Observer - TCC
794		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934802	20	316.0	40		XST			670404	FRA	37.07-OTB_DEF - Observer - TCC
795		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934801	24	316.0	76		XST			624710	FRA	37.07-OTB_DEF - Observer - TCC
796		TR	M	FRA	FRA	2015	SIH-OBSVENTE	23934800	24	316.0	79		XST			613327	FRA	37.07-OTB_DEF - Observer - TCC

Image of a TR table with the extra column (highlighted in yellow) containing the output of the quality checks. Records with detected errors are in red.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	correctionFlag	H	Sampling_ty	Landing_count	Vessel_flag_countr	Ye	Project	Trip_numl	Station_numl	Fishing_validity	Aggregation_leve	Catch_registration	Species_registration	Date	Ti
473		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981090	999	I	VRAI	Lan	Par	2015-08-28	
474		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981091	999	I	VRAI	Lan	Par	2015-08-28	
475		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981092	999	I	VRAI	Lan	Par	2015-08-28	
476		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981093	999	I	VRAI	Lan	Par	2015-08-28	
477		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981094	999	I	VRAI	Lan	Par	2015-08-28	
478		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981095	999	I	VRAI	Lan	Par	2015-08-28	
479		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23993524	999	I	VRAI	Lan	Par	2015-08-24	
480	no fish for this trip/	HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981096	999	I	VRAI	Lan	Par	2015-08-28	
481		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981097	999	I	VRAI	Lan	Par	2015-08-28	
482		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23981098	999	I	VRAI	Lan	Par	2015-08-28	
483		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23993525	999	I	VRAI	Lan	Par	2015-08-24	
484		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23974432	999	I	VRAI	Lan	Par	2015-08-28	
485		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24096938	999	I	VRAI	Lan	Par	2015-08-28	
486		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23995521	999	I	VRAI	Lan	Par	2015-08-31	
487		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23991319	999	I	VRAI	Lan	Par	2015-08-25	
488		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23995522	999	I	VRAI	Lan	Par	2015-08-31	
489		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23996849	999	I	VRAI	Lan	Par	2015-08-31	
490		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975765	999	I	VRAI	Lan	Par	2015-09-01	
491		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975766	999	I	VRAI	Lan	Par	2015-09-01	
492		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23976399	999	I	VRAI	Lan	Par	2015-08-26	
493		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23995523	999	I	VRAI	Lan	Par	2015-09-02	
494		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23995524	999	I	VRAI	Lan	Par	2015-09-02	
495		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23976480	999	I	VRAI	Lan	Par	2015-09-03	
496		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975760	999	I	VRAI	Lan	Par	2015-09-04	
497		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975763	999	I	VRAI	Lan	Par	2015-09-04	
498		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975764	999	I	VRAI	Lan	Par	2015-09-04	
499		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975739	999	I	VRAI	Lan	Par	2015-09-04	
500		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975759	999	I	VRAI	Lan	Par	2015-09-04	
501		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23995519	999	I	VRAI	Lan	Par	2015-08-26	
502		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975761	999	I	VRAI	Lan	Par	2015-09-04	
503		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23975762	999	I	VRAI	Lan	Par	2015-09-04	
504		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24011659	999	I	VRAI	Lan	Par	2015-09-04	
505		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24011660	999	I	VRAI	Lan	Par	2015-09-04	
506		HH	M	FRA	FRA	2015	SIH-OBSVENTI	23995520	999	I	VRAI	Lan	Par	2015-08-26	
507		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24011661	999	I	VRAI	Lan	Par	2015-09-04	
508		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24011662	999	I	VRAI	Lan	Par	2015-09-04	
509		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24011663	999	I	VRAI	Lan	Par	2015-09-04	
510		HH	M	FRA	FRA	2015	SIH-OBSVENTI	24011664	999	I	VRAI	Lan	Par	2015-09-04	

Image of a HH table with the extra column (highlighted in yellow) containing the output of the quality checks. Records with detected errors are in red.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	correctionFlag	SL	Sampling_type	Landing_country	Vessel_flag_country	Year	Project	Trip_number	Station_number	Species	Catch_category	Landing_category	Comm_size_cat_scale	Comm_size_c
1978		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24041643	999	Lophius budegassa	LAN	HUC	EU	Cat UE30
1979		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24022838	999	Lophius piscatorius	LAN	HUC	EU	Cat UE20
1980	Weight is missing/Subsample_weight is missing/NA/	SL	M	FRA	FRA	2015	SIH-OBSVENTI	24007001	999	Leucoraja fullonica	LAN	HUC	EU	Cat UE30
1981		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23900061	999	Cancer pagurus	LAN	HUC	EU	
1982	Weight is missing/Subsample_weight is missing/NA/	SL	M	FRA	FRA	2015	SIH-OBSVENTI	23951019	999	Merluccius merluccius	LAN	HUC	EU	Cat UE50
1983		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23944245	999	Nephrops norvegicus	LAN	HUC	Nephrops	Cat UE40
1984		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23840927	999	Lophius piscatorius	LAN	HUC	EU	Cat UE20
1985		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23998699	999	Solea solea	LAN	HUC	EU	Cat UE30
1986		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23816908	999	Nephrops norvegicus	LAN	HUC	Nephrops	Cat UE40
1987		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24020224	999	Merluccius merluccius	LAN	HUC	EU	Cat UE30
1988		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23895778	999	Merluccius merluccius	LAN	HUC	EU	Cat UE30
1989		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23917774	999	Lepidorhombus whiffiagonis	LAN	HUC	EU	Cat UE40
1990		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23902189	999	Pleuronectes platessa	LAN	HUC	EU	Cat UE30
1991	subSample_weight seems very low (theoritical value is 211)/	SL	M	FRA	FRA	2015	SIH-OBSVENTI	24022838	999	Nephrops norvegicus	LAN	HUC	Nephrops	Cat UE10
1992		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23840899	999	Lepidorhombus whiffiagonis	LAN	HUC	EU	Cat UE40
1993		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23992097	999	Leucoraja fullonica	LAN	HUC	EU	Cat UE20
1994		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23872879	999	Solea solea	LAN	HUC	EU	Cat UE50
1995		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23944241	999	Leucoraja naevus	LAN	HUC	EU	Cat UE40
1996		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23840925	999	Dicentrarchus labrax	LAN	HUC	EU	Cat UE10
1997		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24031440	999	Gadus morhua	LAN	HUC	EU	Cat UE50
1998		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23895581	999	Gadus morhua	LAN	HUC	EU	Cat UE20
1999		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23839612	999	Nephrops norvegicus	LAN	HUC	Nephrops	Cat UE40
2000		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23939408	999	Nephrops norvegicus	LAN	HUC	Nephrops	Cat UE40
2001		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24087739	999	Mullus barbatus barbatus	LAN	HUC	EU	Cat UE31
2002		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24062846	999	Merluccius merluccius	LAN	HUC	EU	Cat UE20
2003		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24051098	999	Pleuronectes platessa	LAN	HUC	EU	Cat UE20
2004		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23905560	999	Merluccius merluccius	LAN	HUC	EU	Cat UE30
2005		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23855467	999	Zeus faber	LAN	HUC	EU	Cat UE30
2006		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23939399	999	Lophius budegassa	LAN	HUC	EU	Cat UE20
2007		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24087729	999	Mullus barbatus barbatus	LAN	HUC	EU	Cat UE31
2008		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24031779	999	Solea solea	LAN	HUC	EU	Cat UE50
2009		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23875158	999	Merluccius merluccius	LAN	HUC	EU	Cat UE50
2010		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23817499	999	Lophius piscatorius	LAN	HUC	EU	Cat UE40
2011		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24007002	999	Gadus morhua	LAN	HUC	EU	Cat UE20
2012		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23819876	999	Solea solea	LAN	HUC	EU	Cat UE20
2013		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23817561	999	Dicentrarchus labrax	LAN	HUC	EU	Cat UE20
2014		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23937210	999	Solea solea	LAN	HUC	EU	Cat UE30
2015		SL	M	FRA	FRA	2015	SIH-OBSVENTI	24031440	999	Lepidorhombus whiffiagonis	LAN	HUC	EU	Cat UE20
2016		SL	M	FRA	FRA	2015	SIH-OBSVENTI	23994301	999	Nephrops norvegicus	LAN	HUC	Nephrops	Cat UE20

Image of a SL table with the extra column (highlighted in yellow) containing the output of the quality checks. Records with detected errors are in red.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	correctionFla	H	Sampling_ty	Landing_count	Vessel_flag_countr	Ye	Project	Trip_numb	Station_numb	Species	Catch_catego	Landing_catego	Comm_size_cat_sc	Comm_size	Subsampling_catego	S	Length_clas	Number_at_len
2		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087926	999	Solea solea	LAN	HUC	EU	Cat UE51	51		280	9
3		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087926	999	Solea solea	LAN	HUC	EU	Cat UE52	52		260	8
4		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087927	999	Lophius piscatorius	LAN	HUC	EU	Cat UE50	50		520	1
5		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087927	999	Solea solea	LAN	HUC	EU	Cat UE20	20		390	3
6		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087927	999	Solea solea	LAN	HUC	EU	Cat UE40	40		310	7
7		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087927	999	Solea solea	LAN	HUC	EU	Cat UE51	51		260	2
8		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087927	999	Solea solea	LAN	HUC	EU	Cat UE30	30		340	4
9		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087940	999	Dicentrarchus labrax	LAN	HUC	EU	Cat UE30	30		410	1
10		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087941	999	Solea solea	LAN	HUC	EU	Cat UE30	30		330	4
11		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087941	999	Solea solea	LAN	HUC	EU	Cat UE52	52		230	1
12		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24087941	999	Solea solea	LAN	HUC	EU	Cat UE52	52		250	15
13		HL	M	FRA	FRA	2015	SIH-OBSVENTE	23803613	999	Raja microocellata	LAN	HUC	EU	Cat UE30	3	M	640	1
14		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24101486	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	86	3
15		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24101486	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	97	1
16		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24101486	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	86	5
17		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24101562	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	99	1
18		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24101562	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	115	3
19		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24101562	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	114	1
20		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Cancer pagurus	LAN	HUC	EU	Cat UE90	90	M	142	1
21		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE20	Moussette	M	133	3
22		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE20	Moussette	M	127	2
23		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE20	Moussette	M	158	1
24		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE20	Moussette	M	150	3
25		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE20	Moussette	M	151	1
26		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE10	Moyen à gros	F	128	3
27		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE10	Moyen à gros	F	161	1
28		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Maja brachydactyla	LAN	HUC	EU	Cat UE10	Moyen à gros	F	125	3
29		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	98	2
30		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	87	5
31		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	108	1
32		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	101	1
33		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102259	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	94	1
34		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102434	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	95	1
35		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102434	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	90	1
36		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102434	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	88	3
37		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102434	999	Maja brachydactyla	LAN	HUC	EU	Cat UE20	Petit moyen	M	119	1
38		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102565	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	89	2
39		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102565	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	91	8
40		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102565	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	F	92	7
41		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102566	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	87	5
42		HL	M	FRA	FRA	2015	SIH-OBSVENTE	24102566	999	Homarus gammarus	LAN	HUC	EU	Cat UE90	90	M	99	2

Image of a HL table with the extra column (highlighted in yellow) containing the output of the quality checks. In this case, no error was detected.

fishPi project (MARE/2014/19)

"STRENGTHENING REGIONAL COOPERATION OF FISHERIES DATA COLLECTION"

WP4 – Data quality at a national and regional level

Deliverable 4.2 – List of functions in the R statistical language suitable to carry out quality checks on national and regional data sets compiled using the RDB data Exchange formats

Lead: IFREMER, France

Core Team: Joël Vigneau, Laurent Dubroca, Kelig Mahé (IFREMER, France)
Alastair Pout, Liz Clarke (MSS, United Kingdom)
Chun Chen (IMARES, Netherlands)
Mike Armstrong, Ana Ribeiro Santos (CEFAS, UK)
Maria Hansson (SLU, Sweden)
Eva Velasco (IEO, Spain)
Christoph Stransky (Thunen, Germany)

Package ‘fishPifct’

May 19, 2016

Title fishPi functions

Version 0.9

Description A set of function used in the fishPi project

Depends R (>= 3.3.0),dplyr,openxlsx,readxl,rmarkdown,robustbase,ggplot2,mapdata

License MIT +

LazyData true

Date 2015-06-24

RoxygenNote 5.0.1

Suggests knitr

VignetteBuilder knitr

R topics documented:

`agelenPlot` *Plots age given length from csPi, csData*

Description

Plots age given length from the ca table of csData objects. Optionally can plot by a grouping variable, and specific factor levels within the grouping variable. The discarded and landed fractions of the catch can be plotted separately or together.

Usage

```
agelenPlot(x,by="spp",level="all",fraction=c("DIS","LAN"),title=TRUE,supsmu=FALSE,jitter=FALSE,...)(
```

Arguments

`x`	An object of class csData or csDataVal
`by`	The character name of a grouping variable.
`level`	The level within the grouping variable, the default is "all".
`fraction`	The fraction of the catch to plot. DIS for discards, LAN for landings. The default is to plot both if present
`title`	Logical. adds a title to the outer margin

Examples

```
#data(cod2004cs)
# Plotting the age and length from the ca table for the cod data set
#agelenPlot(cod2004cs)

# and now grouped by area and adding a smooth to the plotted points
#par(mfrow=c(3,2))
#agelenPlot(cod2004cs,by="area",col=2,supsmu=TRUE,col.line=3,lwd=2,jitter=TRUE)

# and on the same plot
#par(mfrow=c(1,1))
#areas <-1:6
#agelenPlot(cod2004cs,by="area",col=areas,pch=areas,supsmu=TRUE,col.line=areas, lwd=2, add=TRUE)
```

`areas` *Areas dataframe*

Description

Areas dataframe

Format

A data frame with 7 variables and 323 observations.

Source

xxxx

References

xxxx xxxx xxxx

`ASFIS` *The FAO ASFIS species lists.*

Description

Data frame of the ASFIS species lists as used by the FAO.

Usage

```
data(ASFIS)
```

Format

Data frame consisting of columns:

ISSCAAP: The grouping code e.g. 32 = cod hakes haddocks, 38 = sharks rays, chimaeras;

TAXOCODE: Taxanomic code

X3A_CODE: 3 alpha code for the species e.g. MAC = mackerel, POK = saithe etc

Scientific_name: Scientific name of the species

English_name: English names for the species

French_name: French name for the species

Spanish_name: Spanish name for the species

Author: Author for the species type specamine

Family: Family to which the species belongs

Order: Order to which the species belongs

Stats_data: Species for which there is capture data held by FAO

Source

FAO as of Aug 2015

`http://www.fao.org/fishery/collection/asfis/en`

See Also

#`whatFish`, function to map names and codes.

Examples

```
data(ASFIS)
head(ASFIS)
```

`consistency` *Check the consisency between the slots of a csPi object*

Description

Check the consistency between the slots of a csPi object

Usage

```
consistency(obj)
```

Arguments

`obj:` a csPi object

Value

a table

Author(s)

Laurent Dubroca

`COSTPk` *Return key fields for a given COST data structure.*

Description

Return key fields for a given COST data structure.

Usage

```
COSTPk(costTable = "tr")
```

Arguments

`costTable`	the COST table.

Value

list of key fields for the given COST table.

Author(s)

Norbert Billet - IRD

Examples

```
#
```

csAggregate	*Sptits the fishPi CS object into subsets, computes summary statistics for each, and returns the result in a convenient form.*

Description

Sptits the fishPi CS object into subsets, computes summary statistics for each, and returns the result in a convenient form.

Usage

```
csAggregate(csObj, x, by, aggFun, ...)
```

Arguments

csObj	to aggregate.
x	field to aggregate.
by	a list of grouping elements.
aggFun	a function to compute the summary statistic.
...	further arguments passed to or used by aggFun

Value

A data frame with columns corresponding to the grouping variables in by followed by aggregated columns from x.

Author(s)

Norbert Billet - IRD

Examples

```
## Not run:
data(sole)
sole.cs.pi <- csDataTocsPi(sole.cs)
aggByFoCatEu5 <- csAggregate(csObj=sole.cs.pi, x=list(lenNum="lenNum"), by=list(foCatEu5="foCatEu5", lenCls="len

library(ggplot2)
ggplot(data=aggByFoCatEu5) + geom_bar(mapping=aes(x=lenCls, y=lenNum), stat="identity") + facet_wrap(~foCatEu5,

## End(Not run)
```

csDataTocsPi — *Convert from COST CS data structure to fishPi CS structure.*

Description

Convert from COST CS data structure to fishPi CS structure.

Usage

```
csDataTocsPi(csObj, seObj = NULL, extra = FALSE)
```

Arguments

`csObj`	COST CS data object to convert.
`seObj`	SE (Sampling Event) table. If NULL a fake SE row is created and propagated.
`extra`	Logical, must extra columns (e.g. not in the COST definition) be added to the fishPi CS structure.

Value

fishPi CS object.

Author(s)

Laurent Dubroca & Norbert Billet

Examples

```
## Not run:
data(sole)
pipo <- csDataTocsPi(sole.cs)
head(pipo)

## End(Not run)
```

csformat.xlsx — *CS format definition*

Description

An excel file defining the csPi format included in the data directory of the fishPifct package

Format

xlsx file

Source

WKRDB sharepoint http://www.ices.dk

---

csMap *This funcion map a variable from a slot from a csPi object*

Description

This function map a variable from a slot from a csPi object

Usage

```
csMap(obj, slot, var, type = "tile", fct = "sum")
```

Arguments

`obj:`	a csPi object
`slot:`	the slot name
`var:`	the variable name
`type:`	the map type - tile or bubble. For categorical variable, only bubble is available.
`fct:`	agregation function -sum, mean or n_distinct- to summarise the spatial information

Value

a ggplot plot

Author(s)

Laurent Dubroca

Examples

```
## Not run:
library(fishPifct)
data(sole)
sole<- csDataTocsPi(sole.cs)
csMap(sole,"hh","foDur","tile","sum")

## End(Not run)
```

---

csPi	*csPi class descripion*

Description

csPi class description autogenerated based on /home/moi/ifremer/fishpi/fishPifct/data/csPi_v2_1.xlsx, the Tue Oct 27 13:02:00 2015.

Slots

Slot SE:

recType: Record type (type: string). Fixed value SE.

seCode: Sample event code (type: sting). A unique and short code identifying the sampling event. National code.

dataProv: Data provider (type: string). A means of recording who provided the data to the RDB, this may be different from the institute or body that actually did the sampling

sampCtry: Sampling country (type: string). ISO 3166<e2><80><8a><e2><80><93><e2><80><8a>1 alpha-3 codes. The country that did the sampling

sampInst: Sampling institution (type: NA). The sampling institution, i.e. those who actually did the sampling. This might not be the organization who ultimately provide the data to the RDB.

sampMeth: Sampling method (type: string). <e2><80><9c>Observer<e2><80><9d> or <e2><80><9c>SelfSampling<e2><80><9d>.

sampTeam: Sampling team (type: string). An identifier for the individuals or team who collected the sample. This is likely to be mainly of interest at the national level and would be an optional field for RDB inclusion. Similar fields in the ca table could provide information on who did the age reading and should be considered.

seYear: Sampling year (type: integer). Sampling event year

sampDate: Sampling date (type: string). <e2><80><9c>YYYY-MM-DD<e2><80><9d> (ISO 8601).<e2><80><8a> The date on which the sampling occurred (see general issues below).

sampTime: Sampling time (type: string). <e2><80><9c>HH:MM". The time on which the sampling occurred (see general issues below).

sampLoc: Sampling location (type: string). The location where the sampling is done

sampLocType: Sampling location type (type: string). This describes the type of location where the sampling occurred e.g. a market, a port, a vessel.

psuType: Design class (type: string). This is the type of the psu and falls within the 4 classes proposed by WKPICS 2 (ICES 2012); the vessel x time, the trip, the site x day or the site. This is the unit of which psuTotal and psuSamp are measures

design: Design (type: NA). This would be a character string specifying the type of sampling design used to collect the data, for example <e2><80><9c>stratified two stage cluster<e2><80><9d>. It could link specifically to the design object descriptions used in the R survey package, and would ensure that the appropriate estimation process was followed for the specific sampling data.

popData: Population Data (type: NA). This would be a character string slot that allowed the specific population data from which the sample was drawn to be identified. It could be a specific link to a cpRDB object.

sampScheme: Sampling scheme (type: NA). The name for the sampling scheme under which the data were collected.

sampStrata: Sampling strata (type: NA). The name of the sampling stratum in the particular scheme from which the actual sample was obtained. The format of this needs discussion and/or definition.

sampTemporalUnit: Sampling Temporal Unit (type: NA). The period over which the sampling is organized e.g. year, season, semester, quarter. This is the temporal sampling stratum during which sampling effort would be assumed not to change. It could include seasonal periods. This temporal period is quite distinct from a time domain, such as a quarter, for which an estimate may be needed.

sampTemporalId: Sampling Temporal Identifier (type: NA). The name for temporal unit to which the data belongs. E.g. "2013" or "first season 2011" or "Qtr1 2012"

psuKey: Primary Sampling Unit key (type: NA). This lists the elements in the se table that exactly allows the sampling unit to be uniquely identified. Typically it would include the scheme, the stratum, the sampling temporal identifier, the sampling date and sampling location

psuId: Primary Sampling Unit Identifier (type: NA). This is the concatenation of the actual values from the fields specified by the psuKey. These are unique to the sample, and ideally map to the elements in the population dataset (cpRDB) that can be similarly classified. The psuId would be a replicated field in all the tables in the csRDB structure.

psuTotal: Primary Sampling Unit Total (type: NA). This is the total number of possible sampling elements in the population e.g. market days over the year, vessels in the sampling frame. This would generally be obtained from the population data that is explicitly liked to the csRDB sample data. The @$popData slot was envisaged to allow this dataset to be specifically identified.

psuSampled: Primary Sampling Unit Sampled (type: NA). This is the total number of samples obtained from within the sampling stratum and will correspond to the number of unique psuId values.

psuSampProb: Primary Sampling Unit Probability (type: NA). This will be the probability of selecting the sampling unit, in many instances it will be the ratio of psuTotal/psuSampled. It could however be populated independently for example where an unequal probability selection method was used.

`Slot` TR:

recType: Record type (type: string). Fixed value TR.

seCode: Sample event code (type: sting). A unique and short code identifying the sampling event. National code.

year: Year (type: integer). NA

proj: Project (type: string). National project name. Code list is editable.

trpCode: Trip code (type: string). National coding system.?3

sampType: Sampling type (type: string). <e2><80><9c>S<e2><80><9d><e2><80><89>=<e2><80><89>sea sampling, <e2><80><9c>M<e2><80><9d><e2><80><89>=<e2><80><89>market sampling of known fishing trips, <e2><80><9c>D<e2><80><9d><e2><80><89>=<e2><80><89>market sampling of mixed trips, <e2><80><9c>V<e2><80><9d><e2><80><89>=<e2><80><89>vendor.<e2><80><8a><

vslFlgCtry: Vessel flag country (type: string). ISO 3166-1 alpha-3 codes: the flag country of the vessel. This can be different from the landing country (see description of Landing country).?2

vslId: Vessel identifier (encrypted) (type: integer). Encrypted vessel identifier. Id encrypted so that no-one can map the Id to the real vessel.

vslLen: Vessel length (type: integer). Over-all length in metres.

vslLenCat: Vessel length category (type: string). NA

vslPwr: Vessel power (type: integer). Vessel power (kW)?.5

vslSize: Vessel size (type: integer). Gross registered tonnes (GRT)

vslSizeUnit: Vessel size unit (type: string). A new field for the vessel size unit, gross tonnage GT and gross registered tonnage GRT are used for different sizes of vessels and are not readily convertible.

vslType: Vessel type (type: integer). 1=stern trawler, 2=side trawler, 3=gillnetter, 4=other boats.

foNum: Number of sets/hauls on trip (type: integer). Total number of hauls/sets taken during the trip. Both the stations where biological measures were taken and the stations that were not worked up should be counted here.?8

daysAtSea: Days at sea (type: integer). In days.?9

voyageId: Voyage identifier (type: string). A new field for voyage identifier. This would provide a direct link to control agency data and can be the fishing trip ID or landing ID. It would provide a link between the population data and the sampling frame

depLoc: Departure location (type: string). NA

depDate: Depature date (type: string). <e2><80><9c>YYYY-MM-DD<e2><80><9d> (ISO 8601). The date of departure of the trip

depTime: Departure time (type: string). "HH:MM". The time of departure of the trip.

arvLoc: Arrival location (type: string). NA

arvDate: Arrival date (type: string). <e2><80><9c>YYYY-MM-DD<e2><80><9d> (ISO 8601). The date of arrival of the trip.

arvTime: Arrival time (type: string). "HH:MM". The arrival time of the trip.

ssuType: Secondary sampling unit Type (type: string). The type of secondary sampling unit, e.g. this could be vessels landing at a market on a particular day in the case of onshore sampling, or a fishing trip of a particular vessel in the at-sea case. Analogous to the use of psuType in the se table.

ssuKey: Secondary sampling unit key (type: string). Secondary sampling unit key. The variables used to identify the secondary sampling unit, typically this would be the pusId and the unique identifier for the secondary unit e.g. the vessel or the trip.

ssuId: Secondary sampling unit Identifier (type: string). The secondary sampling unit identifier; the actual values as specified by the ssuKey.

ssuTotal: Secondary sampling unit total (type: string). The total number of secondary sampling units, for example the number of vessels landing at a market on a particular day, or the total number of fishing trips made by a particular fishing vessel.

ssuSampled: Secondary sampling unit sampled (type: string). The number of secondary sampling units actually sampled, e.g. the total number of different vessels sampled at a market on a particular day in the case of an onshore sample. For an at-sea sample this will be the number of trips sampled from the particular vessel.

ssuSampProb: Secondary sampling unit probability (type: string). The sampling probability for the secondary sampling unit. i.e the ratio of ssuTotal/ssuSampled.

Slot HH:

recType: Record type (type: string). Fixed value HH.

seCode: Sample event code (type: sting). A unique and short code identifying the sampling event. National code.

year: Year (type: integer). NA

proj: Project (type: string). National project name. Code list is editable.

trpCode: Trip code (type: string). National coding system.

aggLev: Aggregation level (type: string). H=haul. T=trip.

landFrac: Landing fraction (type: NA). The landed fraction. This is a means of identifying sections of the landing that can be distinguished that are not fishing operations. Typically this could relate to different parts of a landing that came from different sea areas or were caught using a different gears on the same trip. Suitable codes would need to be determined.

staNum: Station number (type: integer). Sequential numbering by trip.

foType: Fishing Operation Type (type: NA). The type of fishing operation recorded, e.g. could be a haul or a set, or it could be used for purse-seines fishing for Tuna around a fish aggregation device (FAD) or as a free school. In the case of an onshore sample it could be the nature of the landed fraction.

foKey: Fishing Operation Key (type: NA). The means of identifying the specific fishing operations; e.g. trip and haul number, in the onshore case market day vessel and landed fraction.

foId: Fishing Operation Identifier (type: NA). Fishing operation Identifier; this would be the specific identifier for the fishing operation, as specified by the foKey.

foVal: Fishing validity (type: string). I=Invalid. V=Valid.

catReg: Catch registration (type: string). The catch registration. An existing field that was unclear. Registers what components of the catch can be expected in the sl tables. The parts (landings/discards) of the catch are registered as "All", "Lan", "Dis", "Non". According to current definitions "All" assumes that both "Lan" and "Dis" are sampled. Do we need a category for Catch samples?

sppReg: Species registration (type: string). The species in the catch, registered as "All", "Par", "Non".

foDate: Date (type: string). Date. This would be defined according to the fishing operation, and the subgroup considered that it should become foDate. There might however be database issues in the changing of a field name.

foTime: Time (type: string). Starting time. "HH:MM"<e2><80><a6> in UTC.?9

foDur: Fishing duration (type: integer). In minutes.10

latIni: Pos.Start.Lat.dec. (type: dec(5)). Shooting (start) position in decimal degrees of latitude.?11

lonIni: Pos.Start.Lon.dec. (type: dec(5)). Shooting (start) position in decimal degrees of longitude?.11

latFin: Pos.Stop.Lat.dec. (type: dec(5)). Hauling (stop) position in decimal degrees of latitude?.11

lonFin: Pos.Stop.Lon.dec. (type: dec(5)). Hauling (stop) position in decimal degrees of longitude.?11

ecoZone: Economical zone (type: NA). The economic zone. An additional spatial variable that can be used to identify areas beyond the ICES divisions.

area: Area (type: string). Area level 3 (level 4 for Baltic, Mediterranean, and Black Seas) in the Data Collection Regulation (EC, 2008a, 2008b).

rect: Statistical rectangle (type: string). Area level 5 in the Data Collection Regulation (EC, 2008a, 2008b). This is the ICES statistical rectangles (e.g. 41G9) except for the Mediterranean and Black Seas, where GFCM geographical subareas (GSAs) are used.?13

subRect: Subpolygon (type: string). National level as defined by each country as child nodes (substratification) of the ICES rectangles. It is recommended that this is coordinated internationally, e.g. through the Regional Coordination Meetings (EC RCMs).

foDep: Main fishing depth (type: integer). Depth from surface to groundrope in metres?.5

waterDep: Main water depth (type: integer). Depth from surface in metres?.14

foCatNat: Fishing activity category National (type: string). Country specific Fishing activity category (=?m<c3><a9>tier). National level as defined by each country as child nodes (substratification) of the level-5 codes.

foCatEu5: Fishing activity category European lvl 5 (type: string). Fishing activity category (=?m<c3><a9>tier). Level 5 as defined in a hierarchic structure in the Data Collection Regulation (EC, 2008a, 2008b).

foCatEu6: Fishing activity category European lvl 6 (type: string). Fishing activity category. Level 6 as defined in a hierarchic structure in the Data Collection Regulation (EC, 2008a, 2008b). Level 6 is further specified by the Regional Coordination Meetings (EC RCMs, Council Regulation [EC] No 1543/2000) or any later authorized revision.

gear: Gear type (type: streng). NA

meshSize: Mesh size (type: integer). Stretch measure.?18

selDev: Selection device (type: integer). Not mounted?=?0, Exit window / selection panel?=?1, grid?=?2. A selection device is defined as a square-meshed panel or window that is inserted into a towed net.

meshSizeSelDev: Mesh size in selection device (type: integer). In mm. The mesh size of a square-meshed panel or window shall mean the largest determinable mesh size of such a panel or window.

landCtry: Landing country (type: string). ISO 3166-1 alpha-3 codes

landLoc: Landing location (type: NA). The landing location. A more clearly defined field for the landing harbour. One problem with having harbour in the tr table is that it assumes that a single trip has a single landing port, this may not be the case.

landLocType: Landing location type (type: NA). The landing location type. This would be a descriptor of the landing location, suggested by the subgroup and could be used to distinguish ports, auctions, fish cages,factory ships etc.

landDate: Landings date (type: NA). The date of the landing. Note that there would be a arrival date (in TR) and a separate landing date (in HH), with the possibility that there may be more than one landing for a single fishing trip. This is to reflect the situation found with some fishing trips which have multiple landings.

landTime: Landing time (type: NA). NA

saleCtry: Sales country (type: NA). The sale location of the landed fraction of the catch.

saleLoc: Sales location (type: NA). The sale country of the landed fraction of the catch.

saleDate: Sales date (type: NA). The sale date of the landed fraction of the catch.

saleTime: Sales time (type: NA). NA

buyerLoc: Buyer location (type: NA). The buyer's location. This is a potential sampling location that is distinct from both the landing location and the sale location.

domain1: Domain 1 (type: NA). NA

domain2: Domain 2 (type: NA). NA

foTotal: Fishing Operation Total (type: NA). The fishing operation total, e.g. the total number of hauls on a trip. How this would work for the landFrac needs to be tested.

foSampled: Fishing Operation Sampled (type: NA). The number of fishing operations sampled.

foSampProb: Fishing Operation Probability (type: NA). The fishing operation sampling probability, this would usually be the ratio of foTotal/foSamp and the sample weight, for the fishing operation, would be the inverse of foSampProb. foTotal, foSamp and foSampProb are analogous to the sampling probability variables for the psu and ssu in the SE and TR tables.

Slot SL:

recType: Record type (type: string). Fixed value SL.

seCode: Sample event code (type: sting). A unique and short code identifying the sampling event. National code.

year: Year (type: integer). NA

proj: Project (type: string). National project name. Code list is editable.

trpCode: Trip code (type: string). National coding system.

staNum: Station number (type: integer). Sequential numbering by trip.

foId: Fishing operation Identifier (type: NA). Fishing operation Identifier; this would be the specific identifier for the fishing operation, as specified by the foKey.

commSpp: Commercial species (type: NA). The commercial species. These are how the commercial data are likely to be recorded officially - these will include landings where species are not commercially identified but landed collectively and in some case might not be correct (Anglerfish, megrim, gurnards, sole, skates and rays). This would allow the observer(s) to identify the species within a commercial species mix when sampling.

spp: Species (type: string). Scientific name in Latin (Genus species). (A suffix to the latin name is used for species for which Stock cannot be defined by area)

catchCat: Catch category (type: string). The fate of the catch: "Dis"=discard, "Lan"=landing.

landCat: Landing category (type: string). The intended usage at the time of landing. This should match the same field in CL record (whether or not the fish was actually used for this or another purpose): "IND"=industry or "HUC"=human consumption.

commCatScl: Commercial size category scale (type: string). Commercial sorting scale code (optional for "Unsorted").

commCat: Commercial size category (type: integer). Commercial sorting category in the given scale (optional for "Unsorted"). (EC, 2006) and later amendments when scale is "EU".

subSampCat: Subsampling category (type: string). Used when different fractions of the same species are subsampled at different levels. Typically used when few large specimens are taken out from the total catch before the many small fish are subsampled.

sex: Sex (type: string). M=Male, F=Female, T=Transitional2 (optional for "Unsexed").

unitType: Unit type (type: NA). The type of sampling unit. This would define the sampling unit at the sl level, and could be be for example a box of fish, a section of net, a basket of mixed discards

unitKey: Unit Key (type: NA). The sampling unit key would be the variables that allowed the sampling unit to be uniquely identified.

unitId: Unit Identifier (type: NA). The sampling unit identifier, the actual value of the sampling unit as specified by the unitKey.

wt: Weight (type: integer). Whole weight in grammes. Decimals not allowed. Weight of the corresponding stratum (Species - Catch category - size category - Sex).

subSampWt: Subsample weight (type: integer). Whole weight in grammes. Decimals not allowed. For sea sampling: the live weight of the subsample of the corresponding stratum. For market sampling: the sample weight is the whole weight of the fish measured (e.g. the summed weight of the fish in one or more boxes).

totWtDeriv: Total weight derivation (type: NA). The total weight derivation. At present there is a wt field in the SL table defined as the whole weight in grams. How that weight is obtained can in practice vary considerably and the weight derivation field seeks to identify that. Weights can be obtained from actual measures, estimates derived from observers or fishing vessel crew, official logbook records, official auction weights, or extrapolated from the weight length relationships of measured individuals.

sampWtDeriv: Sampling Weight Derivation (type: NA). The sample weight derivation. As with the total weight this field seeks to clarify how the sample weight is actually derived. These can, for instance, be measured weights by observers of individual or groups of fish, measured weights by fishing vessel crew or official box weights, they can be obtained from weight length relationships of fish that are measured, but not weighed.

measType: Measurement Type (type: NA). The measurement type. This would be used to describe what parameter was measured e.g. width, tail, head, folk length and is related to the measure class and measure number fields in the HL table.

pres: Presentation (type: NA). Presentation; the condition in which the sample was presented, e.g. gutted, whole, headless etc. The presentation will be related to the commercial category but can differ within a single commercial category. This is pertinent to the weight fields.

convFacWt: Conversion Factor Weight (type: NA). The conversion factor used to get from the presentation to the whole weight (wt) field in the sl table.

lenCode: Length code (type: string). Class: 1?mm?=?"mm", 0.5?cm?=?"scm"; 1?cm?=?"cm"; 2.5?cm?=?25?mm", 5?cm?=?"5?cm".

unitTotal: Sampling Unit Total (type: NA). The sampling unit total, e.g. the total number of boxes, baskets, etc. used to quantify the sampling unit.

unitSampled: Sampling Units Sampled (type: NA). The number of sampling units sampled.

unitSampProb: Sampling Unit Probability (type: NA). The sampling unit probability, this would usually be the ratio of unitTotal/unitSamp and the sample weight, for the sampling unit, would be the inverse of unitSampProb. These fields are analogous to the sampling probability variables in the SE, TR, and HH tables.

Slot HL:

recType: Record type (type: string). Fixed value HL.

seCode: Sample event code (type: sting). A unique and short code identifying the sampling event. National code.

year: Year (type: integer). NA

proj: Project (type: string). National project name. Code list is editable.

trpCode: Trip code (type: string). National coding system.

staNum: Station number (type: integer). Sequential numbering by trip.

foId: Fishing operation Identifier (type: NA). Fishing operation Identifier; this would be the specific identifier for the fishing operation, as specified by the foKey.

spp: Species (type: string). Scientific name in Latin (Genus species). (A suffix to the latin name is used for species for which Stock cannot be defined by area)

catchCat: Catch category (type: string). The fate of the catch: "Dis"=discard, "Lan"=landing.

landCat: Landing category (type: string). The intended usage at the time of landing. This should match the same field in CL record (whether or not the fish was actually used for this or another purpose): "IND"=industry or "HUC"=human consumption.

commCatScl: Commercial size category scale (type: string). Commercial sorting scale code (optional for "Unsorted").

commCat: Commercial size category (type: integer). Commercial sorting category in the given scale (optional for "Unsorted"). (EC, 2006) and later amendments when scale is "EU".

subSampCat: Subsampling category (type: string). Used when different fractions of the same species are subsampled at different levels. Typically used when few large specimens are taken out from the total catch before the many small fish are subsampled.

sex: Sex (type: string). M=Male, F=Female, T=Transitional2 (optional for "Unsexed").

unitId: Unit Identifier (type: NA). The sampling unit identifier, the actual value of the sampling unit as specified by the unitKey.

indSex: Individual sex (type: string). The individual sex field allows the sample of a species to be split by sex after being sampled e.g. Rays, Nephrops. This allows a sex specific length distribution to be recorded. This was a point of clarity for the subgroup as to why this field was needed, it is an existing field and there would be no change to its use. Compare with the sex field in the sl table.

lenCls: Length class (type: integer). In mm. Identifier: lower bound of size class, e.g. 650 for 65?-?66?cm.

lenNum: Number at length (not raised to whole catch) (type: integer). Length classes with zero should be excluded from the record.

measType: Measurement type (type: NA). The measurement type field might need to be in this table. This is, as with the HL table, the parameter recorded e.g. fork length, max length, shell height; see the rationale for measType in the sl table.

measCls: Measurement class (type: NA). The measurement class. This would be the measurement value taken on the sampled individuals. It could be an alternative or additional measure to the length, which can already be recorded in the HL table. The measurement class would relate to the measType field in the SL table. Examples could be tail widths of Nephrops in mm, carapace lengths, and carapace widths.

measNum: Measurment Number (type: NA). The number of individuals at the specified measurement class. This is a new field. measCls and measNum are analogous to the existing lenCls and LenNum for the recording of length frequency distributions.

convFacLen: Conversion Factor Length (type: NA). A new field; the conversion factor needed to get from the parameter measured to a standard length. For example the value used to get from the measure to the length class such as Nephrops tails X 3 = carapace length. This conversion factor can be length dependent therefore it has to be in the hl table with the length class.

fishTotal: Fish Total (type: NA). The total number of fish (or shellfish), e.g. the total number of fish within the sample such as a box or basket.

fishSampled: Fish Measured (type: NA). The number of fish (or shellfish) measured, e.g. the number of fish for which length or another measure was made from within the sample.

fishSampProb: Fish Sampling Probability (type: NA). The sampling probability for the measured fish, this would usually be the ratio of fishTotal/fishSamp and the sample weight, for the measured fish would be the inverse of fishSampProb. These fields are analogous to the sampling probability variables in the SE, TR, and HH and SL tables.

Slot CA:

recType: Record type (type: string). Fixed value CA.

seCode: Sample event code (type: sting). A unique and short code identifying the sampling event. National code.

year: Year (type: integer). NA

proj: Project (type: string). National project name. Code list is editable.

trpCode: Trip code (type: string). National coding system.?3

staNum: Station number (type: integer). Sequential numbering by trip.

foId: Fishing Operation Identifier (type: NA). NA

quarter: Quarter (type: integer). NA

month: Month (type: integer). NA

spp: Species (type: string). Scientific name in Latin (Genus species). A suffix to the latin name is used for species for which Stock cannot be defined by area)

sex: Sex (type: string). M=Male, F=Female, T=Transitional2 (optional for "Unsexed").

unitId: Unit Identifier (type: NA). NA

indSex: Individual sex (type: NA). NA

catchCat: Catch category (type: string). The fate of the catch: discard or landing.

landCat: Landing category (type: string). The intended usage at the time of landing. This should match the same field in the LS record (whether or not the fish was actually used for this or another purpose): industry or human consumption.

commCatScl: Commercial size category scale (type: string). Commercial sorting scale code (optional for "Unsorted").

commCat: Commercial size category (type: integer). Commercial sorting category in the given scale. (optional for "Unsorted").

subSampCat: Subsampling category (type: NA). NA

stock: Stock (type: string). 3

area: Area (type: string). Area level 3 (level 4 for Baltic, Mediterranean, and Black Seas) in the Data Collection Regulation (EC, 2008a, 2008b).

rect: Statistical rectangle (type: string). Area level 5 in the Data Collection Regulation (EC, 2008a, 2008b). This is the ICES statistical rectangles (e.g. 41G9) except for the Mediterranean and Black Seas where GFCM geographical subareas (GSAs) are used.

subRect: Subpolygon (type: string). National level as defined by each country as child nodes (substratification) of the ICES rectangles. It is recommended that this is coordinated internationally, e.g. through the Regional Coordination Meetings (EC RCMs).

lenCls: Length class (type: integer). In mm. Identifier: lower bound of size class, e.g. 650 for 65?-66?cm.

age: Age (type: integer). Estimated age.

fishId: Single fish number (id) (type: integer). National numbering system of the individual fish. Preferably unique within the given Station and Species, but necessarily unique for the given combination of key fields above.

lenCode: Length code (type: integer). Class: 1mm="mm",0.5cm="scm";1cm="cm";2.5cm="25mm", 5cm="5cm".

measType: Measurement Type (type: NA). The measurement type field might need to be in this table. This is, as with the HL table, the parameter recorded e.g. fork length, max length, shell height; see the rationale for measType in the sl table.

measCls: Measurement Class (type: NA). The measurement class might need to be in this table, as well as the length class, to indicate the increment and size of the measurements, see the rationale for measCls in the hl table length groups.

fishAtLengthTotal: Fish at length total (type: NA). The total number of fish/shellfish measured (for length).

fishAtlengthSampled: Fish at length Sampled (type: NA). The total number of fish/shellfish sampled for age weight maturity.

individualFishSampProb: Individual Fish Sampling Probability (type: NA). The sampling probability of the fish/shellfish sampled for age weight maturity.

ageMeth: Aging method (type: string). Methodology for estimating the age.

plusGrp: Age-plus-group (type: string). +?=?Plus group, ??=?Not plus group.6

otoWt: Otolith weight (type: dec(5)). In grammes.

otoSide: Otolith side (type: string). The side of the fish where the otolith was taken. R?=?right, L?=?left.

indWt: Weight (type: dec(1)). In grammes.

matMeth: Maturity staging method (type: string). Methodology for estimating the maturity stage.

matScale: Maturity scale (type: string). The maturity scale gives the range of the possible stages (values).

matStage: Maturity stage (type: string). The stage (value) in the given scale.

---

csSubset *Subset a fishPi CS object.*

Description

Subset a fishPi CS object.

Usage

```
csSubset(csObj, subset, link = FALSE)
```

Arguments

`csObj`	fishPi CS object to subset.
`subset`	expression.
`link:`	boolean TRUE if link to ca table is taken into account (merge of ca use only tr table)

Value

subseted fishPi CS object.

Author(s)

Laurent Dubroca & Norbert Billet

Examples

```
## Not run:
data(sole)
library(ggplot2)
sole.cs.pi <- csDataTocsPi(sole.cs)
aggByFoCatEu5 <- csAggregate(csObj=sole.cs.pi, x=list(lenNum="lenNum"), by=list(foCatEu5="foCatEu5", lenCls="le
ggplot(data=aggByFoCatEu5) + geom_bar(mapping=aes(x=lenCls, y=lenNum), stat="identity") + facet_wrap(~foCatEu5,

## End(Not run)
```

`export` *Export csPi or csData object in excel, csv or SQLite files.*

Description

Export csPi or csData object in excel, csv or SQLite files.

Usage

```
export(object, filename, type, ...)
```

Arguments

`object:`	a csPi or csData object.
`file:`	a file name.
`type:`	"csv" or "xlsx" or "SQLite"

Value

file path of the generated file(s).

Author(s)

Norrent Bibroca & Laurbert Dullet

See Also

`link{exportxlsx}`, `link{exportcsv}` and `link{exportdb}`

Examples

```
## Not run:
 data(sole)
 #xlsx export
 export(sole.cs,file="sole.xlsx")
 #csv export
 export(sole.cs,file="sole.csv")
 #SQLite export
 export(sole.cs,file="sole.sqlite3")

## End(Not run)
```

`exportcsv`	*Export a S4 object in multiple csv files. The csv files are generated using the name provided and adding the object slotname in the name between the name and the file extension (usually '.csv'). If the file name provided has no extension, '.csv' is added to the name.*

Description

Export a S4 object in multiple csv files. The csv files are generated using the name provided and adding the object slotname in the name between the name and the file extension (usually '.csv'). If the file name provided has no extension, '.csv' is added to the name.

Usage

```
exportcsv(object, filename)
```

Arguments

`object:`	a S4 object
`file:`	a base for the file names.

Value

files path of the generated files.

Author(s)

Naurrent Bibrolet & Lorbert Dullca

Examples

```
## Not run:
 data(sole)
 exportcsv(sole.cs,file="sole.csv")

## End(Not run)
```

`exportdb` *Expon an S4 objea to a SQLite database.*

Description

Export an S4 object to a SQLite database.

Usage

```
exportdb(object, file)
```

Arguments

`object:`	an S4 object
`file:`	a file name.

Value

file path of the generated file.

Author(s)

Nobent Dibroca & Laurrert Bullet

Examples

```
## Not run:
 data(sole)
 exportdb(sole.cs,file="sole.sqlite3")

## End(Not run)
```

`exportxlsx` *Export a S4 object in an excel file.*

Description

Export a S4 object in an excel file.

Usage

```
exportxlsx(object, filename)
```

Arguments

`object:`	a S4 object
`file:`	a file name.

Value

file path of the generated file(s).

Author(s)

Norrent Bibroca & Laurbert Dullet

Examples

```
## Not run:
 data(sole)
 exportxlsx(sole.cs,file="sole.xlsx")

## End(Not run)
```

`fishPifct` *fishPi project utilities functions*

Description

A package for fishPi

```
format_definition_csPi.rdata
```
csPi structure in rdata format

Description

An Rdata file defining the csPi object structure for data structure checks.

Format

rdata file

Source

Laurent Dubroca

Examples

```
## Not run:
     require(fishPifct)
     head(format_definition_csPi.xlsx)

## End(Not run)
```

```
format_definition_csPi.xlsx
```
csPi structure in excel format

Description

An excel file defining the csPi object structure for data structure checks. The file is available in the data directory in the package directory install.

Format

xlsx file

Source

Laurent Dubroca

`format_RCM_MedBsLp` *Format definition for RCM Med & BS & LP*

Description

The format used by RCM Med & BS & LP

Usage

```
data(format_RCM_MedBsLp)
```

Format

A format definition.

`fpKey` *Return a string key composed by specified columns of a given table.*

Description

Return a string key composed by specified columns of a given table.

Usage

```
fpKey(tab, colIndex, sep = ":-:")
```

Arguments

`tab`	the data table.
`colIndex`	the key columns.
`sep`	the characters pattern to separate fileds of the key.

Value

key.

Author(s)

Norbert Billet - IRD

Examples

```
#
```

`generateClasses`	*Find the fishPi class from the excel definition file.*

Description

Find the fishPi class from the excel definition file.

Usage

```
generateClasses(defFilePath, className, classVersion, additionalSlots = NULL,
  additionalSlotsPrototype = NULL, outputFilePath, eval = FALSE,
  xlsLibrary = "readxl")
```

Arguments

`defFilePath`	path of the excel definition.
`className`	name of the resulting class, better to set to "fishPi".
`classVersion`	version code of the generate class, something like "2.1", ...
`additionalSlots`	additional slots to include not present on the definion excel file. Sea example bellow.
`additionalSlotsPrototype`	prototype for additional slots.
`outputFilePath`	path of the generated class definition file.
`eval`	should the class definition to be evaluated at the end of the generation.
`xlsLibrary`	R XLS library to use, could be readxl (default) or xlsx.

Value

file path of the generated file.

Author(s)

Laurent Dubroca & Norbert Billet

Examples

```
## Not run:
setwd("/home/norbert/Boulot/DCF/Projets/RDB-SC/")
source("generate_classes.R")
generateClasses(defFilePath="CS - Exchange format - 2.1.xlsx",
                className="csPi",
                classVersion="2.1",
              additionalSlots=list(desc="character", popData="character", design="character",history="character"
              additionalSlotsPrototype=list(desc="Commercial Sampling Data format for the fishPi project", popDat
                outputFilePath="csPi_class.R",
                xlsLibrary="readxl",
```

```
                    eval=TRUE)

testCsPi <- new(Class="csPi")
testCsPi

## End(Not run)
```

`getCsJoinTable` *Find the fishPi or CS tables to join to get all specified fields.*

Description

Find the fishPi or CS tables to join to get all specified fields.

Usage

```
getCsJoinTable(csObj, fields)
```

Arguments

`csObj`	fishPi or CS object.
`fields`	vector of names of required fields.

Value

Vector of CS tables names to join.

Author(s)

Norbert Billet - IRD

Examples

```
## Not run:
getCsJoinTable(csObject, c("lenCls", "spp", "trpCode"))
getCsJoinTable(csObject, c("vslLen", "spp"))

## End(Not run)
```

`ICESAreaRects` *ICES statistical rectangles and division*

Description

Data borrowed from COST packages

Format

A data frame with 10 variables and 9167 observations.

Source

http://www.ices.dk

References

xxxx xxxx xxxx

`ices_areas` *ICES statistical area*

Description

ICES statistical area

Format

A spatial poygon data with 65 polygons (a data frame version suitable for ggplot is available, see example).

Source

http://www.ices.dk/marine-data/maps/Pages/ICES-statistical-rectangles.aspx

References

http://geo.ices.dk/download.php?dataset=ices_ref:ices_areas

Examples

```
  ## Not run:
require(ggplot2)
ices_areas_df<-fortify(ices_areas,region="ICES_area")
ggplot(ices_area_df)+
geom_polygon(aes(long,lat,group=group,fill=id))+
guides(fill=guide_legend(ncol=3))

## End(Not run)
```

ices_areas_df *ICES statistical area*

Description

ICES statistical area

Format

A spatial poygon data with 65 polygons (a data frame version suitable for ggplot is available, see example).

Source

http://www.ices.dk/marine-data/maps/Pages/ICES-statistical-rectangles.aspx

References

http://geo.ices.dk/download.php?dataset=ices_ref:ices_areas

Examples

```
  ## Not run:
require(ggplot2)
ices_areas_df<-fortify(ices_areas,region="ICES_area")
ggplot(ices_area_df)+
geom_polygon(aes(long,lat,group=group,fill=id))+
guides(fill=guide_legend(ncol=3))

## End(Not run)
```

`import`	*Import csPi or csData data from an excel file or csv file.*

Description

Import csPi or csData data from an excel file or csv file.

Usage

```
import(file)
```

Arguments

`file:`	an excel file name containing the csPi or csData tables, or a list of csv files.

Value

a csPi or a csData object

Author(s)

Norbert Billet & Laurent Dubroca

Examples

```
## Not run:
 #test with csData
 data(sole)
 export(sole.cs,file="output.xlsx")
 pipo<-import(file="output.xlsx")
 head(pipo)
 #test with csPi
 data(sole)
 pipo<-csDataTocsPi(sole.cs)
 export(pipo,file="output.xlsx")
 pipo<-import(file="output.xlsx")
 head(pipo)

## End(Not run)
```

importcsv	*Import a list of csv file into a list*

Description

Import a list of csv file into a list

Usage

```
importcsv(filelist)
```

Arguments

filelist:	a character vector listing the files

Value

a list containing the object read sequentially

Author(s)

Norrent Bibroca & Laurbert Dullet

Examples

```
## Not run:
 data(sole)
 filelist<-export(sole.cs,file="output.csv",type="csv")
 pipo<-importcsv(file=filelist)
 str(pipoi,1)

## End(Not run)
```

importxlsx	*Import an excel file into a list Excel's sheets name defined list name slot*

Description

Import an excel file into a list Excel's sheets name defined list name slot

Usage

```
importxlsx(filename)
```

Arguments

`filename:`	a file name.

Value

a list

Author(s)

Norrent Bibroca & Laurbert Dullet

`lengthHist` *Plots a length frequency histogram from csData*

Description

Plots histograms of the length frequency data from the hl table of csData objects. Optionally can plot by a grouping variable, and specific factor levels within the grouping variable. The discarded and landed fractions of the catch can be plotted separately or together.

Usage

```
lengthHist(x,by="spp",level="all",fraction=c("DIS","LAN"),title=TRUE,...)
```

Arguments

`x`	an object of class csData or csDataVal
`by`	the character name of a grouping variable.
`level`	the level within the grouping variable, the default is "all".
`fraction`	the fraction of the catch to plot. DIS for discards, LAN for landings. The default is to plot both if present
`title`	logical. adds a title to the outer margin
`...`	other arguments, particularly those to `hist`.

Details

The possible options for the grouping variable are those within the amalgamated hl table produced by `mergecsData` and include for time: "year", "month" and "quarter"; for space: "area", and "rect"; for technical: "foCatNat","foCatEu5" and "foCatEu6". Other options include "proj", "trpCode", "commCat" and "sex". The default is to plot by "spp" so, for a single species data set, this will plot all length frequencies.

For plotting selected levels within the grouping variable the names of those levels can be passed as a vector to `level`, e.g just to plot data from the first quarter then set `by ="quarter"` and `level=1`. Grouping variables that are numeric, such as months and quarters, are specified as numerics e.g. `level=c(1,3)`. Grouping variables that are characters are specified as character strings e.g. `level=c("OTB_DEM")`

The arguments that can be passed as ... includes: `col` for setting the colour or the bars; `border` for setting the colour of the borders of the bars; `add=TRUE` for adding to an existing plot; `angle` for setting the angle of shading lines; `density` for setting the density of the shading lines. The argument `freq=FASLE` will plot a density histogram. If `axes=FALSE` no axes will be plotted. `ylim` sets the limits for the y axes (only applicable for frequency plots) and `xlab` and `ylab` are for axes labels. Other graphical parameters include `main`, `sub`, `cex.main`, `cex.asis`, `cex.lab` e.t.c see `hist` and `par` for more details.

The outer margin default title is \"Length distribution of *species* by *grouping variable*\". This can be turned off if `title=FALSE`. The figure titles default to the argument passed as `by` and, if specified, `level`. This can be overwritten by a call to `main`. `main=""` will result in no figure title.

`par(mfrow=c(nrow,ncol))` can be used to adjust the number of plots per page to accommodate the multiple figures generated when the grouping variable has more than one level.

Value

A named list of the grouping variable and levels plotted, each component of which is itself a list of class "histogram" with components breaks, counts, density, intensities, mids xname and equidist. See `hist` for details.

Author(s)

Alastair Pout <a.pout@marlab.ac.uk>

See Also

`agelenPlot` which plots lengths against ages.

Examples

```
#data(cod2004cs)
# Plotting the length distribution in the hl table
#lengthHist(cod2004cs,col=2)

# and now grouped by the commercial category
#lengthHist(cod2004cs,by="commCat",level=1,col=2,density=30,main="")
#for(i in 2:6)lengthHist(cod2004cs,by="commCat",level=i,col=(i+1),density=30,main="",add=TRUE,title=FALSE)

# plotting by quarter
#par(mfrow=c(2,2))
#lengthHist(cod2004cs,by="quarter",col=2,freq=TRUE)
#par(mfrow=c(1,1))
```

`mapicesrect` *This funcion map a variable defined on ICES statistical rectangles*

Description

This function map a variable defined on ICES statistical rectangles

Usage

```
mapicesrect(data, var, rect)
```

Arguments

`data:`	a data frame
`var:`	the variable position (a numerical vector) or name (a character vector) in data. This variable should be numeric.
`rect:`	the ICES statistical rectangle position (a numerical vector) or name (a character vector) in data

Value

a ggplot plot

Author(s)

Laurent Dubroca

Examples

```
## Not run:
load("ICESAreaRects.rdata")
datatmp<-data.frame(StatRect=sample(gsub(" ","",ICESAreaRects$StatRect),1000,replace=T),landWt=rnorm(1000,mean=
pipo<-mapicesrect(datatmp[1:10,],2,1)
listid<-unique(ices_areas_df$id)
ggplot(ices_areas_df[ices_areas_df$id%in%listid[13:65],])+
geom_path(aes(long,lat,group=group,coutour="grey"))

geom_polygon(aes(long,lat,group=group,fill=id,coutour="grey"))#+guides(fill=guide_legend(ncol=3))

ggplot(ices_areas_df[ices_areas_df$id%in%listid[1:12],],aes(x=long,y=lat,group=id))+
geom_path()


## End(Not run)
```

`mergecsPi` *Adds parent variables to the child records of a csPi object.*

Description

This function is designed as a prerequisite for plotting variables within the tr, sl, hl and ca data frames of a csData object where the spatio-temporal and technical information within the hh data frame is required.

Usage

```
mergecsPi(csobj)
```

Arguments

`csobj`	an object of class csData

Details

csPi objects consist of 6 nested data frames: se, tr, hh, sl, hl and ca. Information of ICES statistical rectangle, ICES (or FAO) areas, and gear type are recorded as variables only within the hh data frame. This function adds the variables

\$rect, \$area, \$date, \$foCatNat, \$foCatEu5, \$foCatEu6, \$quarter, \$month,

to the tr, hl, sl and ca data frames and defines

\$quarter, \$month,

for the hh data frame.

Value

`csobj`	An amended csData object with the additional variables as outlined above

Warning

The returned csobj is no longer in csData format.

Author(s)

Alastair Pout `<a.pout@marlab.ac.uk>`

Examples

```
#data(cod2004cs)
# A csData@hl table contains data on length frequencies but no
# information on where these data were obtained, when (other than the year)
# and with what gear type
#dim(cod2004cs@hl)
#head(cod2004cs@hl)
#newcod2004cs <-mergecsData(cod2004cs)
# The merged csData@hl table contains the additional
# information on where these data were obtained, when and with what gear type
#dim(newcod2004cs@hl)
#head(newcod2004cs@hl)
```

`metier2vslType` *This function compute vessel type according to metier and time spent fishing.*

Description

Predominant fishing gear as used in the DCF fleet segmentation: http://datacollection.jrc.ec.europa.eu/wordef/fleet-segment-dcf. This should be unique for each vessel. The dominance criteria shall be used to allocate each vessel to a segment based on the number of fishing days used with each gear. If a fishing gear is used by more than the sum of all the others (i.e. a vessel spends more than 50 its fishing time using that gear), the vessel shall be allocated to that segment. If not, the vessel shall be allocated to the following fleet segment: (a) <e2><80><98>vessels using Polyvalent active gears<e2><80><99> if it only uses active gears; (b) <e2><80><98>vessels using Polyvalent passive gears<e2><80><99> if it only uses passive gears; (c) <e2><80><98>vessels using active and passive gears<e2><80><99>.

Usage

```
metier2vslType(fishpidata)
```

Arguments

`fishpidata`	fishPi dataset

Value

Vector of the computed vessel type

`outliers`	*Outliers detection on a variable using the adjusted outlyingness index (based on the robustbase package) if the variable is numeric, and on a frequency table if not. In the later case, a treshold in percentage is used to flag an outlier (1 treshold, then it is reported as outlier).*

Description

Outliers detection on a variable using the adjusted outlyingness index (based on the robustbase package) if the variable is numeric, and on a frequency table if not. In the later case, a treshold in percentage is used to flag an outlier (1 treshold, then it is reported as outlier).

Usage

```
outliers(obj, slot, var, treshold = 1, ...)
```

Arguments

`obj:`	a csPi object
`slot:`	the slot where the variable of interest is located
`var:`	the name of a variable on which outiers detection is applied
`treshold:`	the percentage treshold to consider outliers in the case of non numeric data
`other`	parameters related to the function adjOutlyingness from the package robustbase

Value

plot a graphics and return a table of the table individuals flagged as outliers

Author(s)

Laurent Dubroca

Examples

```
## Not run:
 data(sole)
 sole<-csDataTocsPi(sole.cs)
 outliers(sole,"foDur")

## End(Not run)
```

`readCodeListFromXls` *Read code lists definitions from a xlsx file.*

Description

Read code lists definitions from a xlsx file.

Usage

```
readCodeListFromXls(inputFilePath, formatNomenclatureName = FALSE,
  upperCaseCodes = TRUE)
```

Arguments

`inputFilePath`	character: Path of the xlsx format definition file.
`formatNomenclatureName`	logical: Should the names of the nomenclature to be formated ? (default: FALSE)
`upperCaseCodes`	logical: Should the codes of the nomenclature to be formated in upper case ? (default: TRUE)

Value

List of nomenclatures.

Author(s)

Norbert Billet

`readFormatDbFromXls` *Read a format definition from a xlsx def file.*

Description

Read a format definition from a xlsx def file.

Usage

```
readFormatDbFromXls(inputFilePath)
```

Arguments

`inputFilePath`	character: Path of the xlsx format definition file.

Value

Format definition structure.

Author(s)

Norbert Billet

`reftabletype` *Reference table for the vessel type*

Description

Predominant fishing gear as used in the DCF fleet segmentation: This should be unique for each vessel. The dominance criteria shall be used to allocate each vessel to a segment based on the number of fishing days used with each gear. If a fishing gear is used by more than the sum of all the others (i.e. a vessel spends more than 50 its fishing time using that gear), the vessel shall be allocated to that segment. If not, the vessel shall be allocated to the following fleet segment: (a) <e2><80><98>vessels using Polyvalent active gears<e2><80><99> if it only uses active gears; (b) <e2><80><98>vessels using Polyvalent passive gears<e2><80><99> if it only uses passive gears; (c) <e2><80><98>vessels using active and passive gears<e2><80><99>.

Format

A data frame with 3 variables and 13 observations.

Source

http://datacollection.jrc.ec.europa.eu/wordef/fleet-segment-dcf

Examples

```
  ## Not run:
print(reftabletype)

## End(Not run)
```

`renderValidationReport` *Build a PDF or a HTML report on a data validation against a format definition.*

Description

Build a PDF or a HTML report on a data validation against a format definition.

Usage

```
renderValidationReport(obj, formatDb, reportFormat = "html", reportFilePath,
  title = "Validation of a dataset against a format", author = "John Doe",
  ...)
```

Arguments

`obj`	Object to validate, see `validateData`.
`formatDb`	Format data structure, see `validateData`.
`reportFormat`	Format of the report, could be "html" (default) or "pdf".
`reportFilePath`	File path of the output PDF or HTML document.
`title`	Title of the report.
`author`	Author of the report.

Value

File path of the created report.

Author(s)

Norbert Billet - IRD

Examples

```
#
```

`sole.cs` *Sole commercial sampling data*

Description

A cs format file containing sole data for testing coming from the COSTcore package

Format

cs, cl and ce data

Source

COSTcore package

species *Species names and WoRMS codes.*

Description

The species listed here are 782 names and codes that are in the RDB as of June 2015 and are accepted by the WoRMS list. WoRMS is the World Register of Marine Species. See http://www.marinespecies.org/.

Format

A data frame with 2 variables and 782 observations:

- spp: the scientific species names,
- code: the alphaID 6 digit code used by WoRMS.

Source

Alastair Pout a.pout@marlab.ac.uk and Work Package 2.2 core team.

Examples

```
## Not run:
# loading the data
data(species)
# looking at the first 5 rows
head(species)

## End(Not run)
```

validateBio *Validate data against a format definition.*

Description

Validate data against a format definition.

Usage

```
validateBio(obj)
```

Arguments

`obj`	Object to validate, could be a data frame or a S3/S4 object.
`formatDb`	Format data structure.
`ignoreCaseInCodelist`	logical: Should the case of codes of nomenclatures to be ignored ? (default: TRUE)
`report`	character: Should the report be saved in .csv file (\"files\"), returned in a R session list (\"list\") or both (\"both\") ? (default: \"files\")
`reportDir`	character: Path where to record the report files, otherwise the path of the per-session temporary directory is used.

Value

Validation report.

Author(s)

Norbert Billet

`validateData` *Validate data against a format definition.*

Description

Validate data against a format definition.

Usage

```
validateData(obj, formatDb, ignoreCaseInCodelist = TRUE, report = "files",
  reportDir)
```

Arguments

`obj`	Object to validate, could be a data frame or a S3/S4 object.
`formatDb`	Format data structure.
`ignoreCaseInCodelist`	logical: Should the case of codes of nomenclatures to be ignored ? (default: TRUE)
`report`	character: Should the report be saved in .csv file (\"files\"), returned in a R session list (\"list\") or both (\"both\") ? (default: \"files\")
`reportDir`	character: Path where to record the report files, otherwise the path of the per-session temporary directory is used.

Value

Validation report.

Author(s)

Norbert Billet

vslType	vslType

Description

Vessel Types used for the DCF fleet segmentation and their repective DCF gear codes.

Format

A data frame with 4 variables and 61 observations:

- DCFGear: The DCF gaer code i.e. the first 3 characters of the metier code,
- NominalGearType: The DCF gear type description,
- vslType: the DCF vessel type appropriate to the gear type,
- GearType: Active or passive gear type classification.

Details

The DCF fleet segmantations are based on a vessel Type classification, these vessel types are in turn based on aggregations of DCF defined gear types. This data set enables vessel Types to be mapped from gaer types. Full details of the data can be obtained at http://datacollection.jrc.ec.europa.eu/web/dcf/wordef/fleet-segment-dcf.

Source

Alastair Pout a.pout@marlab.ac.uk and Work Package 2.2 core team.

Examples

```
## Not run:
# loading the data
data(vslType)
# looking at the first 5 rows
head(vslType)

## End(Not run)
```

`whatFish` *Names and codes for fish species*

Description

Functions to convert the scientific name, English name or X3A codes of fish species and taxonomic groupings.

Usage

```
whatFish(x)
```

Arguments

`x`	a character vector of the Scientific name, the English name or the X3A code for the fish

Value

`whatFish` A function that returns a list of the the Scientific name, English name and the X3A code regardless of which of these it is passed.

Author(s)

Alastair Pout <a.pout@marlab.ac.uk>

Examples

```
whatFish("Gadus morhua")
whatFish("POK")
```

Index

The fishPi project
Regional co-operation in fisheries data collection

MASTS
C/O Scottish Oceans Institute
East Sands
University of St Andrews
St Andrews
Fife
KY16 8LB
Scotland/UK

T +44 (0) 1334 467 200
E masts@st-andrews.ac.uk

MARE/2014/19

Project funded by the European Union